







# NKNK: a New Essential Motif in the C-Terminal Domain of HIV-1 Group M Integrases

Marine Kanja,<sup>a,b</sup>  Pierre Cappy,<sup>a,b</sup> Nicolas Levy,<sup>c</sup> Oyndamola Oladosu,<sup>c</sup> Sylvie Schmidt,<sup>d</sup> Paola Rossolillo,<sup>a,b</sup> Flore Winter,<sup>a,b</sup> Romain Gasser,<sup>a,b</sup>  Christiane Moog,<sup>d</sup> Marc Ruff,<sup>c</sup>  Matteo Negroni,<sup>a,b</sup>  Daniela Lener<sup>a,b</sup>

<sup>a</sup>Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, Strasbourg, France

<sup>b</sup>Interdisciplinary Thematic Institute (ITI) InnoVec, Strasbourg, France

<sup>c</sup>Chromatin Stability and DNA Mobility, Department of Integrated Structural Biology, IGBMC, Strasbourg University, CNRS, INSERM, Illkirch, France

<sup>d</sup>Molecular Immuno-Rheumatology Laboratory, UMR1109, FMTS, Université de Strasbourg, INSERM, Institut de Virologie, Strasbourg, France

**ABSTRACT** Using coevolution network interference based on comparison of two phylogenetically distantly related isolates, one from the main group M and the other from the minor group O of HIV-1, we identify, in the C-terminal domain (CTD) of integrase, a new functional motif constituted by four noncontiguous amino acids (N<sub>222</sub>K<sub>240</sub>N<sub>254</sub>K<sub>273</sub>). Mutating the lysines abolishes integration through decreased 3' processing and inefficient nuclear import of reverse-transcribed genomes. Solution of the crystal structures of wild-type (wt) and mutated CTDs shows that the motif generates a positive surface potential that is important for integration. The number of charges in the motif appears more crucial than their position within the motif. Indeed, the positions of the K's could be permutated or additional K's could be inserted in the motif, generally without affecting integration *per se*. Despite this potential genetic flexibility, the NKNK arrangement is strictly conserved in natural sequences, indicative of an effective purifying selection exerted at steps other than integration. Accordingly, reverse transcription was reduced even in the mutants that retained wt integration levels, indicating that specifically the wt sequence is optimal for carrying out the multiple functions that integrase exerts. We propose that the existence of several amino acid arrangements within the motif, with comparable efficiencies of integration *per se*, might have constituted an asset for the acquisition of additional functions during viral evolution.

**IMPORTANCE** Intensive studies of HIV-1 have revealed its extraordinary ability to adapt to environmental and immunological challenges, an ability that is also at the basis of antiviral treatment escape. Here, by deconvoluting the different roles of the viral integrase in the various steps of the infectious cycle, we report how the existence of alternative equally efficient structural arrangements for carrying out one function opens up the possibility of adapting to the optimization of further functionalities exerted by the same protein. Such a property provides an asset to increase the efficiency of the infectious process. On the other hand, though, the identification of this new motif provides a potential target for interfering simultaneously with multiple functions of the protein.

**KEYWORDS** integrase, evolution, human immunodeficiency virus, phylogenetic groups

Integration of reverse-transcribed viral genomes into the genome of the infected cell is a peculiar feature of the replication strategy of retroviruses, carried out by the viral enzyme integrase (IN) in a two-step reaction. In human immunodeficiency virus type 1 (HIV-1), after the achievement of DNA synthesis in the cytoplasm of the infected cell, it first catalyzes the removal of a conserved GT dinucleotide from the 3' ends of the viral

**Citation** Kanja M, Cappy P, Levy N, Oladosu O, Schmidt S, Rossolillo P, Winter F, Gasser R, Moog C, Ruff M, Negroni M, Lener D. 2020. NKNK: a new essential motif in the C-terminal domain of HIV-1 group M integrases. *J Virol* 94:e01035-20. <https://doi.org/10.1128/JVI.01035-20>.

**Editor** Viviana Simon, Icahn School of Medicine at Mount Sinai

**Copyright** © 2020 American Society for Microbiology. All Rights Reserved.

Address correspondence to Matteo Negroni, [m.negroni@ibmc-cnrs.unistra.fr](mailto:m.negroni@ibmc-cnrs.unistra.fr), or Daniela Lener, [d.lener@ibmc-cnrs.unistra.fr](mailto:d.lener@ibmc-cnrs.unistra.fr).

**Received** 26 May 2020

**Accepted** 17 July 2020

**Accepted manuscript posted online** 29 July 2020

**Published** 29 September 2020

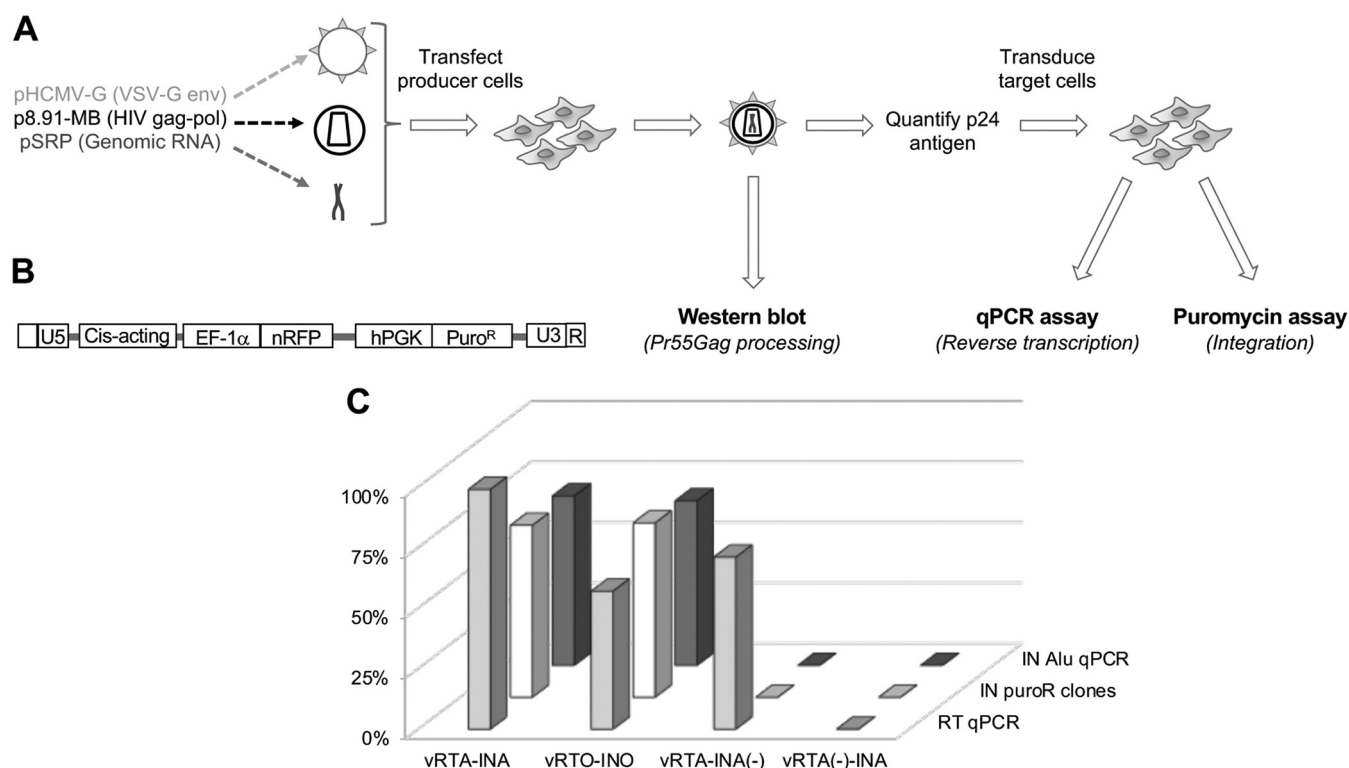
DNA (3' processing), leaving CA<sub>-OH</sub> 3' ends bound to the active site. Subsequently, once the viral DNA has been imported into the nucleus, the reactive CA<sub>-OH</sub> 3' ends attack the cellular DNA, leading to the generation of the provirus (1, 2).

Besides this enzymatic function, HIV-1 IN is involved, through noncatalytic activities, in several other steps of the viral replication cycle. As a component of the Gag-Pol polyprotein precursor, it participates in Gag-Pol dimerization, essential for the autoactivation of the viral protease and, consequently, for viral particle maturation (3–5). During capsid morphogenesis, it is involved in the recruitment of the genomic RNA inside the core of the viral particle (6). As a mature protein, it interacts with the viral polymerase (reverse transcriptase [RT]) to optimize reverse transcription of the viral genome (7–9). Through the interaction with the cellular protein LEDGF/p75, it targets actively transcribed genes as sites for integration (10). Finally, as a component of the preintegration complex (PIC), IN is also involved in nuclear import of the reverse transcription product, a peculiar feature of lentiviruses that allows the infection of nondividing cells.

This ability relies on the capacity of the virus to enter the nucleus via an active passage through the nuclear pore complex (NPC) (11, 12). Several lines of evidence have indicated that the capsid (CA) protein is crucial for nuclear entry (13, 14), through its interaction with several nucleoporins (Nups) forming the NPC (Nup358, Nup153, Nup98) (15–17) and with transportin-3 (18, 19). Nevertheless, several studies have indicated that IN has karyophilic properties. Namely, it contains a basic bipartite nuclear localization signal (NLS) (20) as well as an atypical NLS (21) and also binds several cellular nuclear import factors. Interactions of IN with importin  $\alpha/\beta$  (22), importin 7 (23), importin  $\alpha 3$  (24), Nup153 (25), Nup62 (26), and transportin-3 (27–29) have been documented. Indeed, the mutation of amino acids, mostly located in the C-terminal domain of the IN, responsible for binding to nuclear import factors, results in noninfectious viruses impaired in nuclear import (23, 24, 28, 30).

The functional form of the HIV-1 integrase is made up of a dimer of dimers, which assemble in highly ordered multimers of these tetramers (31, 32). Three domains, connected by flexible linkers, constitute HIV-1 IN: the N-terminal domain (NTD), the catalytic core domain (CCD), and the C-terminal domain (CTD) (33, 34). While the NTD is mostly involved in protein multimerization (35, 36), the CCD is mostly responsible for catalysis and for binding to the viral and cellular DNA, as well as to the cellular cofactor LEDGF (37–40). Finally, the CTD is involved in DNA binding during integration (41, 42), in protein multimerization (35), in the interaction with the reverse transcriptase (7, 9), and in the recruitment of the viral genomic RNA (gRNA) in the viral core (6). Overall, the intrinsic flexibility of the protein, the multiple steps required to achieve integration, and the multimeric nature of the integration complex make the involvement of the different parts of the protein in the various functions of the integrase very complex and still not fully elucidated.

In addition, the multiple tasks that the IN must accomplish during the infectious cycle and the complexity of its supramolecular structures are expected to impose functional constraints that ultimately may limit its genetic diversity. Retention of functionality despite sequence variation strongly relies on covariation, inside or outside the mutated protein. When an initial mutation negatively alters the protein functionality, compensatory mutations can restore it, at least partially. Therefore, the sequences of homologous proteins in different HIV variants are the result of independent evolution pathways, with independent covariation networks specifically generated by each pathway. Chimeric genes between variants of a given protein can perturb such networks and result in the production of nonfunctional proteins. This information can then be exploited to probe the existence of functional motifs in proteins. For considerably divergent viruses, such as those derived from independent zoonotic transmissions, this approach can be particularly powerful. This is the case for HIV-1 groups M and O, which derive from simian viruses infecting chimpanzees and gorillas, respectively. Here, we exploit the natural genetic diversity existing between these groups to generate chimeric integrases. A detailed characterization of the individual amino acids

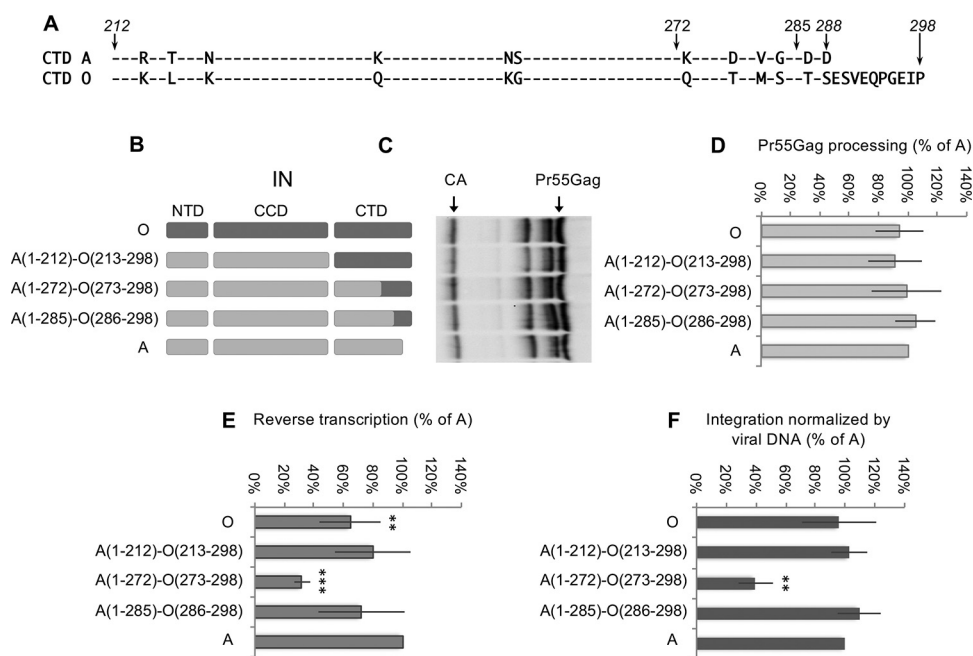


**FIG 1** Outline of the experimental system. (A) Workflow used to evaluate Pr55Gag processing, reverse transcription, and integration in our experimental system. Vesicular stomatitis virus (VSV)-pseudotyped HIV-1-derived vectors, produced by triple transfection, were used to transduce HEK 293T cells. Upon integration, the proviral DNA allows growth of cellular clones in the presence of puromycin. For multiplicities of infection lower than 1, the number of clones obtained is directly proportional to the number of integration events. (B) Schematic representation of the viral genomic RNA contained in the viral vectors, transcribed from pSRP (see panel A and Materials and Methods). R, U5, and U3, viral sequences constituting the LTR; cis-acting, viral sequences required for RNA packaging and reverse transcription; EF-1 $\alpha$  and hPGK, internal human promoters driving the expression of the nuclear RFP (nRFP) and of the puromycin *N*-acetyltransferase vector. The results are the average values of three independent experiments.

that differ in the nonfunctional chimeras then led to the identification and functional characterization of a new motif, in the CTD of HIV-1 group M integrase, essential for viral integration.

## RESULTS

**Analysis of intergroup M/O chimeras in the CTD of IN.** The functionality of the integrases studied in this work was evaluated by following the protocol outlined in Fig. 1A and B and detailed in Materials and Methods. For this, we replaced the original RT and IN sequences of p8.91-MB (see Materials and Methods) by those of either one isolate of HIV-1 group M, subtype A2, referred herein as isolate A, or one isolate of HIV-1 group O, referred herein as isolate O. The resulting viral vectors were named vRTA-INA and vRTO-INO, respectively. With these vectors, we estimated the functionality of the integrases by measuring the efficiency of generation of proviral DNAs. Since the number of proviral DNAs generated for each sample is dependent not only on the levels of functionality of the integrase but also on the amount of total viral DNA generated after reverse transcription, we estimated the amount of total viral DNA generated by each sample by quantitative PCR (qPCR) as described in Materials and Methods. In parallel, we measured the amount of proviral DNA generated either by the puromycin assay or by the Alu qPCR assay as described in Materials and Methods (see "Evaluation of integration by puromycin assay"). The amount of proviral DNA divided by that of total viral DNA provides an estimate of the efficiency of integration. Comparable efficiencies of integration were measured with the two vectors, irrespective of whether the estimation was done using the puromycin assay ( $71\% \pm 13\%$  and  $72\% \pm 24\%$ ,

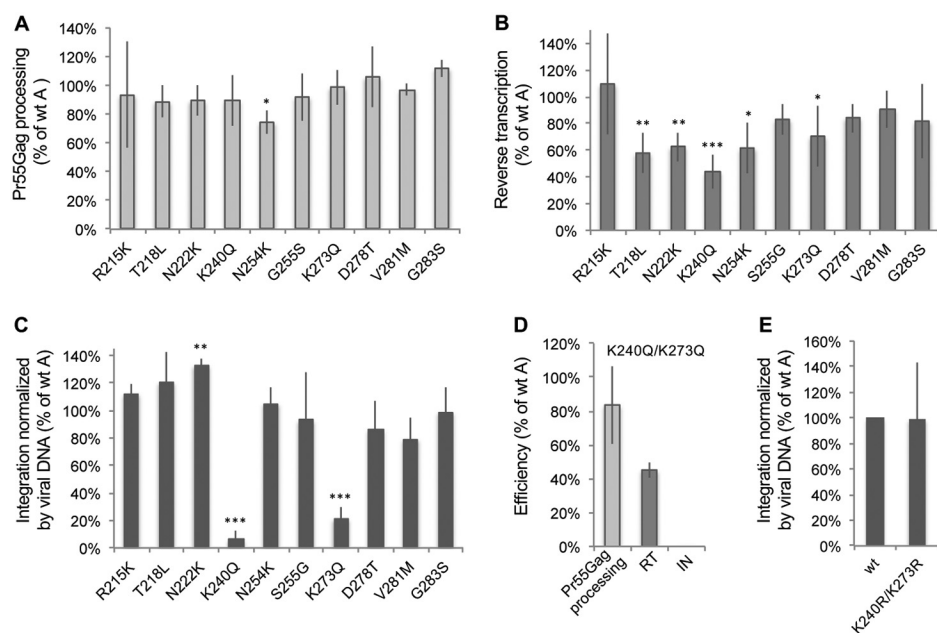


**FIG 2** Functionality of chimeric integrases. (A) Alignment of CTD sequences from isolates A and O, used in this study. The numbers in *italic* on the left and right of the alignment indicate the beginning and the end (in amino acids) of the CTD, respectively. Only amino acids divergent between the two sequences are indicated by letters. The arrows and numbers above the alignment indicate the last position that, in the chimeras, was concordant with the sequence of isolate A. (B) Schematic representation of the integrases studied. Integrase from isolate O is represented at the top of the panel in dark gray; integrase from isolate A is represented at the bottom of the panel in light gray. The genetic origins of the portions of the chimeras are indicated by the colors representing the reference isolates A and O. (C) Representative Western blot obtained with an anti-CA mouse monoclonal antibody. (D) Efficiency of processing of the Pr55Gag precursor, estimated by the amount of CA compared to the amount of Pr55Gag precursors detected by Western blotting (as in panel B). The results are expressed as a function of the reference wt INA, set at 100%. (E) Efficiency of reverse transcription (detection of the U5-Psi junction by qPCR) expressed as a function of the reference wt INA. (F) Efficiency of integration calculated by the puromycin assay, normalized by the amount of total viral DNA (estimated by qPCR), expressed as a function of the reference wt INA. Error bars represent standard deviations. The results given in panels C to E are the average of three independent experiments. \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ , for comparison to wt INA.

respectively, of the level of the reference vector v8.91-MB) or the Alu qPCR assay ( $70\% \pm 14\%$  and  $68\% \pm 15\%$ , respectively) (Fig. 1C). Throughout this work, the efficiency of integration was always evaluated by the puromycin resistance assay normalized by the amount of total DNA. Control vectors in which the catalytic activity either of the integrase or of the reverse transcriptase were abolished in vRTA-INA by the introduction of the D116A mutation in IN, or of the D110N and D185N mutations in RT (43, 44), gave the expected results (Fig. 1C).

We chose to probe the existence of functional motifs in the CTD of integrase, because this domain is involved in several noncatalytic functions of the protein. The CTD of the IN of isolate O (INO) used in this study is 10 amino acids longer (amino acids 212 to 298) than that of the IN of isolate A (INA) (amino acids 212 to 288) (Fig. 2A). We constructed three chimeras between isolates A and O, named after the position in amino acids from the beginning of the IN-coding region, where the sequence shifts from that of one isolate to that of the other (Fig. 2B). Chimera A(1-212)-O(213-298) is constituted by INA with the entire CTD from INO; chimera A(1-285)-O(286-298) is INA with the additional 10 amino acids of INO at the C-terminal end plus the two most C-terminal different amino acids. Finally, as the region between position 212 and 288 differs in 12 amino acids, chimera A(1-272)-O(273-298) was constructed in such a way as to split the 12 different amino acids into two groups of 6.

We first performed Western blot analysis (Fig. 2C) on viral particles to monitor the degree of proteolytic processing of the Gag precursor (Pr55Gag), since incomplete processing would result in immature viral particles, affecting infectivity. No significant



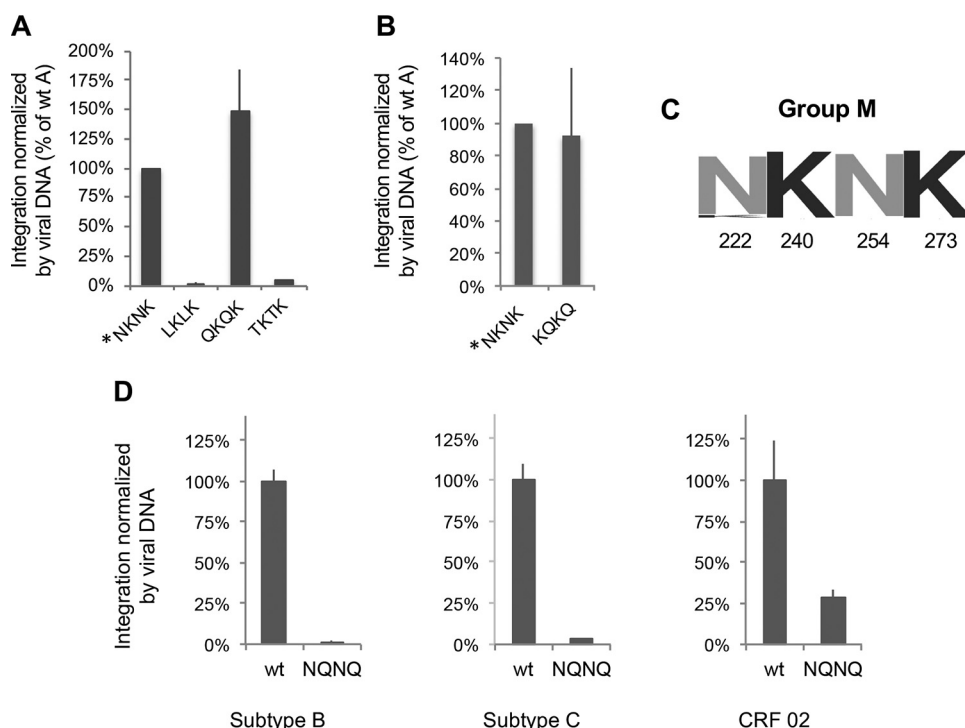
**FIG 3** Functionality of INA with mutated CTD. (A to C) Efficiency of processing of the Pr55Gag precursor (A), of reverse transcription (B), and of normalized efficiency of integration (C). (D) Efficiency of processing of the Pr55Gag precursor, of reverse transcription, and of normalized efficiency of integration for the K240Q K273Q mutant. (E) Efficiency of integration of the K240R K273R mutant and of wt INA (set at 100%). Error bars represent standard deviations. In all panels, the results are the average of three independent experiments. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ , for comparison to wt INA.

differences in Pr55Gag processing were observed between isolate O and chimeric constructs in comparison to isolate A (Fig. 2D). We then evaluated the efficiency of reverse transcription (measuring the amount of viral DNA produced by qPCR) and of integration (as described above). Only chimera A(1–272)-O(273–298) exhibited significant defects in both reverse transcription and integration (Fig. 2E and F), suggesting that a covariation network, present between positions 212 and 285, was broken in this chimera. Since in these experiments the IN is expressed from p8.91-MB and not from the genomic RNA, it can be ruled out that the phenotypes observed are due to an effect of the mutations on the genomic RNA, as for example on the process of splicing, as has been previously described for some mutants of the CTD of IN (45).

**Characterization of IN CTD.** In order to evaluate the individual contribution of the 10 amino acids differing between positions 212 and 285 (Fig. 2A), each residue in INA was individually replaced by those of INO and the 10 point mutants were tested for processing of Pr55Gag, reverse transcription, and generation of integrated proviruses.

Except for the N254K mutant, no significant difference in level of Pr55Gag proteolytic processing was observed between the mutants and parental vector A (Fig. 3A). The effect on reverse transcription was an overall reduction of efficiency for most of the mutants, with a residual efficiency between 45 and 90% of that of the parental vector A (Fig. 3B). Concerning integration efficiency, instead, the majority of the mutants did not show a significant decrease, except for the K240Q and K273Q mutants, for which integration was dramatically impaired (Fig. 3C). This suggests a specific implication of these two residues in the integration process. When the two mutations were combined (K240Q K273Q mutant), although the level of reverse transcription remained above 40% of that of wild-type (wt) INA, integration dropped to undetectable levels (Fig. 3D).

To discriminate between the role of the charge of K<sub>240</sub> and K<sub>273</sub> from that of their possible acetylation, we replaced both residues with two R's (K240R K273R mutant). The level of integration of this mutant was comparable to that of wt INA (Fig. 3E), indicating that the presence of a positive charge and not acetylation at these positions was important for integration efficiency. However, these mutations reduced by half both Pr55Gag proteolytic processing and reverse transcription (data available upon request).

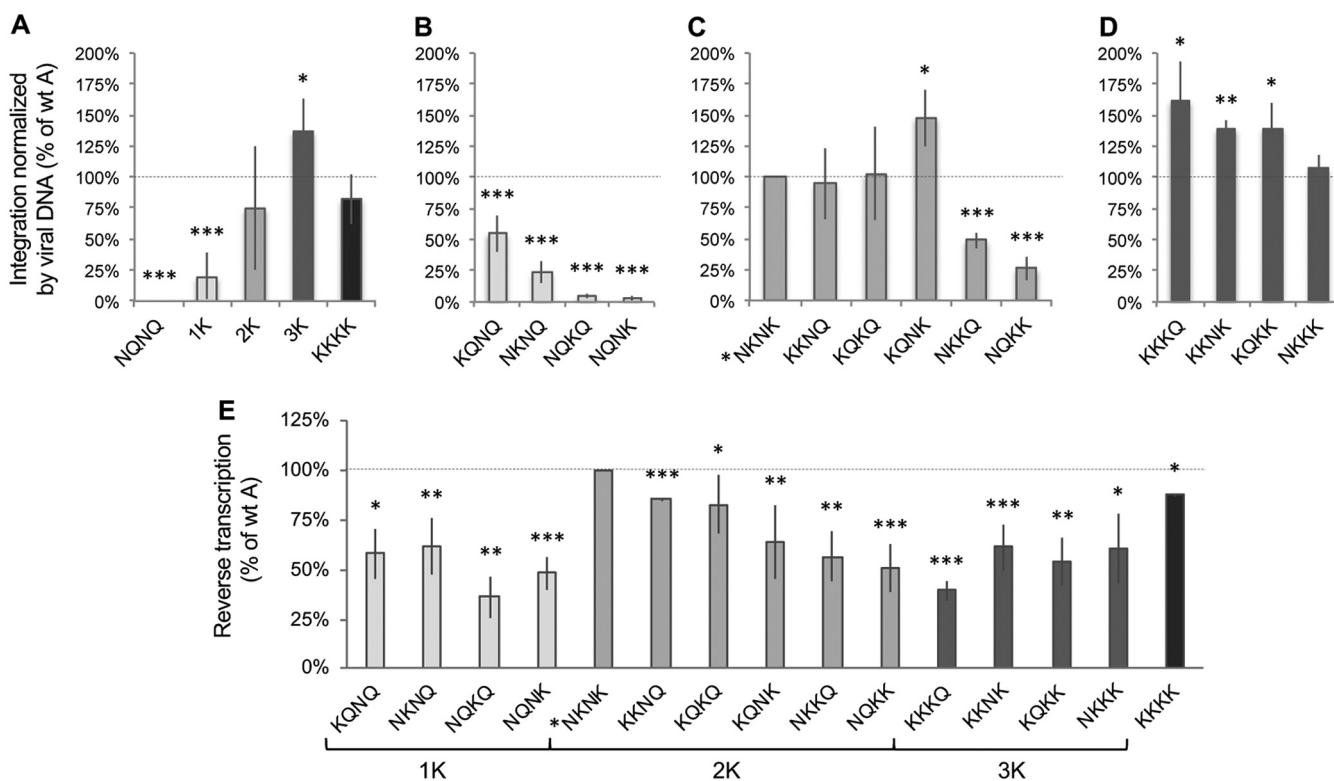


**FIG 4** Definition of the NKNK motif and of its importance in the most widespread phylogenetic groups of HIV-1. (A) Efficiency of integration of various mutants of the N residues of the NKNK motif and of the wt enzyme (\*NKNK, set at 100%). (B) Efficiency of integration of the mutants carrying the sequences of isolate O at positions 222, 240, 254, and 273 ( $K_{222}Q_{240}K_{254}Q_{273}$ ; KQKQ in the figure) and of the wt enzyme (\*NKNK, set at 100%). (C) Conservation logo of the sequence at positions 222, 240, 254, and 273 in HIV-1 group M integrases. (D) Efficiency of integration of the K240Q K273Q double mutant (NQNQ) of an isolate of subtype B, one of subtype C, and one from CRF02, compared to the corresponding wt integrases, set as the reference at 100%. In all vectors, the RT sequence had the same phylogenetic origin as IN and was replaced using the MluI-BspEI cassette in p8.91-MB, as described in Materials and Methods. Error bars represent standard deviations (standard deviations for the wt integrases of each subtype were calculated with respect to reference wt INA, used as the control). The results are the average of three independent experiments.

In both mutants showing a marked defect in integration (K240Q and K273Q mutants), a K (positively charged polar side chain) was replaced with a Q (noncharged polar side chain), the amino acid present in isolate O at the corresponding positions. Conversely, in isolate O, two K's are present in positions where a polar noncharged amino acid (N in both cases) is present in isolate A (positions 222 and 254) (Fig. 2A). Therefore, in order to evaluate if the two noncharged polar amino acids (N) present in isolate A at positions 222 and 254 are also essential, we replaced them with a nonpolar amino acid like leucine (LKLK mutant) and, in parallel, with a noncharged polar residue, Q (QKQK mutant). While in the LKLK mutant the efficiency of integration dropped to almost undetectable levels, in the QKQK mutant it was comparable to that of the wt enzyme, suggesting that the presence of a polar residue at these positions is essential (Fig. 4A). To understand whether the polar nature of the amino acids at positions 222 and 254 is enough to retain functionality, the N's were replaced with two threonines, which are polar but do not have the amide group of asparagine. In this case (TKTK mutant), integration dropped to undetectable levels (Fig. 4A), indicating that not only is the polarity important but also the functional group carried by the amino acid. Therefore, the biochemical features of all four residues identified are important.

Finally, we wondered whether the residues present at positions 222, 240, 254, and 273 could be interchanged between isolates O and A. Therefore, we generated the N222K K240Q N254K K273Q quadruple mutant of isolate A IN (called the KQKQ mutant for simplicity). Remarkably, the integration efficiency of this mutant was not significantly different from that of wt INA (Fig. 4B), indicating the existence of a functional link between these four positions.



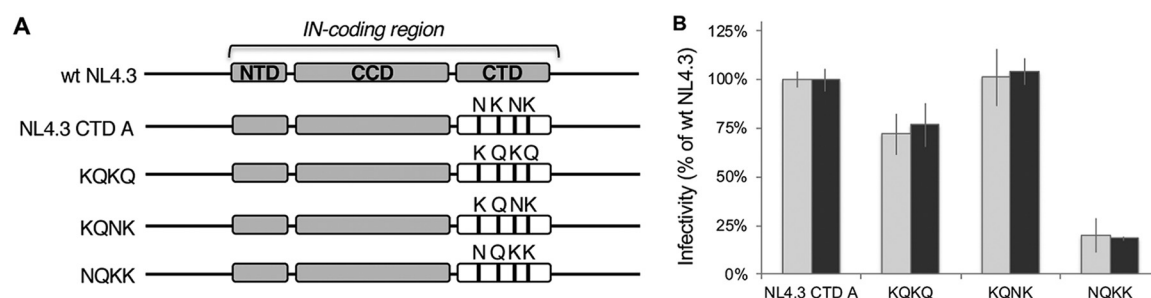


**FIG 5** Importance of the number and position of the K residues in the  $N_{222}K_{240}N_{254}K_{273}$  motif of the CTD. For clarity, only the four letters of the amino acids of the motif are represented for each mutant, omitting the positions; the first letter indicates the residue at position 222, the second at position 240, the third at position 254, and the fourth at position 273. (A) Efficiency of integration, normalized by the amount of viral DNA, for the IN mutants grouped by the number of K's present at positions 222, 240, 254, and 273. The composition in amino acids at the four positions of the motif is given only for the groups represented by one mutant: the group with no K and the one with four K's. (B to D) Efficiency of integration of the individual mutants containing one (B), two (C), or three (D) K's in the motif. In panel C, the motif corresponding to the sequence of wt INA (reference set at 100%) is indicated by an asterisk. Error bars represent standard deviations. The results are the average of four independent experiments. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ , for comparison to wt INA. (E) Efficiency of reverse transcription for the mutants containing one, two, three, or four K's in the motif. The motif of wt INA is indicated by an asterisk. The colors of the bars reflect the colors used in panels A, B, C, and D. The results are the average of at least three independent experiments: \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ , for comparison to wt INA.

The alignment of HIV-1 IN sequences reveals a strong conservation of amino acids  $N_{222}K_{240}N_{254}K_{273}$  in group M (Fig. 4C). To confirm the need for  $K_{240}$  and  $K_{273}$ , observed in isolate A, also for other isolates of group M, we introduced the K240Q K273Q double mutation (NQNNQ mutant) in integrases from three other primary isolates of group M (Fig. 4D). In all cases, a dramatic drop in integration was observed with respect to the corresponding wt integrases, confirming the results obtained with isolate A. The importance of the two K's in the motif was therefore confirmed in isolates from the most widespread HIV-1 group M subtypes in the epidemics, subtypes A, B, C, and CRF02, which are responsible for 79% of the HIV-1 infections worldwide (46).

The possibility of permutating the four amino acids at positions 222, 240, 254, and 273 indicates a functional relationship between these residues. Collectively, based on the data described above and on the sequence conservation present in group M, hereafter we refer to the four non-contiguous  $N_{222}K_{240}N_{254}K_{273}$  amino acids as the "NKNK" motif.

**Importance of the lysines in the NKNK motif.** To understand to what extent the number and positions of the K's in the motif influence IN functionality, we generated a series of mutants based on the replacement of the amino acids present in isolate A with those of isolate O. We thus tested all possible variants (Fig. 5A) containing either one K (four mutants) (Fig. 5B), two K's (five mutants plus the wt) (Fig. 5C), three K's (four mutants) (Fig. 5D), or four K's (one mutant, KKKK) (Fig. 5A) at any of the positions in the motif. The presence of a single K led on average to a drop to 20% of integration with



**FIG 6** Importance of the NKNK motif in replication-competent viruses. (A) Scheme of the portion coding for the integrase in various viruses. Drawn in gray are the parts derived from the NL4.3 sequences; those drawn in white are from isolate A. The vertical black bars indicate positions 222, 240, 254, and 273 from left to right; the amino acid found for each mutant at each of these four positions is indicated above the bars. (B) Infectivity of the viruses shown in panel A (except for wt NL4.3, which is used as the reference, set at 100%). The results are given in gray for CEM-SS cells and in black for TZM-bl cells. Error bars represent standard deviations with respect to the reference, wt pNL4-3. The results are the average of two independent experiments.

respect to wt INA, whereas when two or more lysines were present in the motif, the levels of integration were close to those of wt INA, ranging from 75% to 137% (Fig. 5A). The mutant with no K's, where the motif sequence has been changed from NKNK to NQNK, confirmed the total loss of integration already observed with this mutant (Fig. 3D). Finally, from the analyses of the different mutants, it appears that the presence of a K in the first position of the motif (position 222) consistently leads to a higher level of integration in all classes of mutants (those with one, two, or three K's). Interestingly, though, position 222 has an N in the wt enzyme.

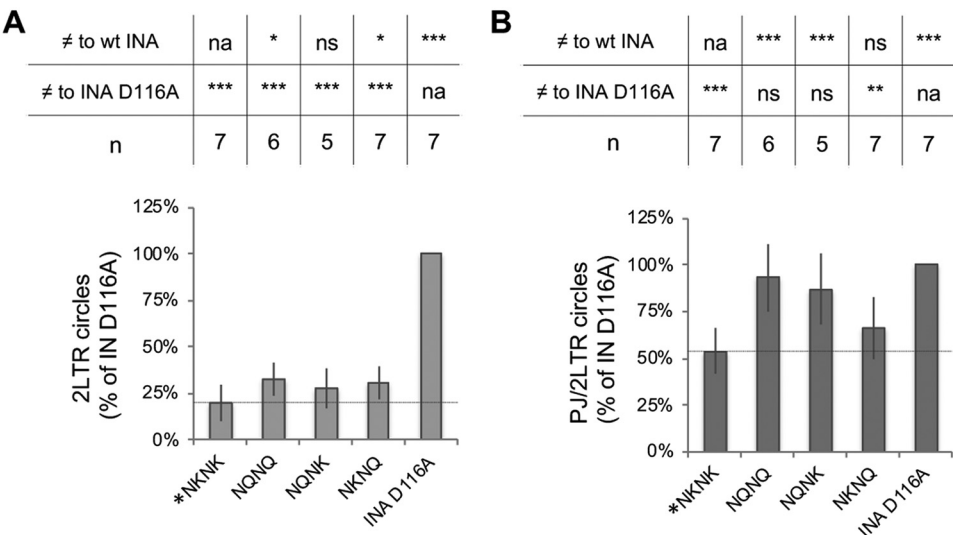
When considering individual mutants within the different classes, we observed a significant decrease in functionality for all the mutants possessing only one K (Fig. 5B). For the mutants containing two K's, three variants were at least as functional as wt INA (NKNK in the figure), while two displayed a significant reduction (Fig. 5C). Finally, all mutants containing three K's were at least as functional as the parental INA (Fig. 5D). Remarkably, the results obtained with the mutants containing three or four K's indicate that the positively charged residues can replace the polar ones, while the reverse is not the case, as shown by the mutants with no K or only one K. Overall, these results indicate that at least two K's are required to have wt levels of integration, even if not all the positions in the motif are equivalent. Instead, all mutants impacted reverse transcription with a reduction ranging from 40% to 80% of that of the wt INA (Fig. 5E).

**The NKNK motif in replication-competent viruses.** To confirm the observations made in the single infection cycle system, some mutants were then tested in a replication-competent system using NL4.3 as the primary virus. Mutants of the class containing two K's in the motif (the number of K's found in circulating viruses) and with a marked phenotype were chosen for this analysis. Besides the wt A sequence, we chose three mutants that either retained integration (KQKQ and KQNK) or exhibited reduced integration (NQKK) (Fig. 5C). To construct the four variants, we replaced the sequence of NL4.3 CTD with that of isolate A, either wt or carrying the KQKQ, KQNK, or NQKK motifs (Fig. 6A).

The infectivity of the virus carrying the whole CTD of INA instead of that of NL4.3 (called NL4.3 CTD A) (Fig. 6A) was comparable to that of wt NL4.3 virus, set as the reference, indicating that the replacement of the whole CTD from NL4.3 with that of isolate A did not impact viral infectivity (Fig. 6B). Regarding the mutants, the results well recaptured the observations made with a single infection cycle (Fig. 5C): the infectivity was maintained for KQKQ and KQNK mutants, while it was markedly decreased with the NQKK motif (Fig. 6B).

**Role of the lysines of the motif in the integration process.** In order to characterize in which steps of the infectious cycle are involved the lysines of the NKNK motif, we evaluated the effect of their mutation in two steps (other than reverse transcription) upstream the integration of the preproviral DNA in the chromosomes of the host cell.





**FIG 7** Amount of 2LTRc/total viral DNA (A) and the ratio of PJ/2LTRc (B) in the mutants deprived of one or both K's of the NKNK motif. The motif corresponding to the sequence of wt INA (reference set at 100%) is indicated by an asterisk in the bar graphs. Error bars represent standard deviations. Above the graphs are given the *P* values for the comparisons of the different samples to wt INA and to the INA D116A integration-deficient mutant (\*, *P* < 0.05; \*\*, *P* < 0.01; \*\*\*, *P* < 0.001). na, not applicable; ns, not significant. The number of independent experiments performed for each sample (*n*) is also given. The amount of 2LTRc for sample X is normalized by the total amount of viral DNA (detection of the late reverse transcription products by quantifying the U5-Psi junction) and then it is expressed as a percentage of the amount detected for the control INA D116A (represented by D), thus giving  $[(2LTRc_X/r_X)/(2LTRc_D/r_D)] \times 100$ , where  $r_X$  and  $r_D$  are the amounts of late reverse transcription products in sample X and control D, respectively, and 2LTRc<sub>X</sub> and 2LTRc<sub>D</sub> are the amounts of 2LTRc in sample X and n control D, respectively.

In particular, by quantifying the two long terminal repeat (LTR) circles (2LTRc), we evaluated nuclear import and, by characterizing the LTR-LTR junctions of 2LTRc, the efficiency of 3' processing, which takes place in the cytoplasm, before nuclear import.

2LTRc are exclusively formed in the nucleus and are, therefore, useful markers for nuclear import of the reverse-transcribed genomes (47). They are generated when the full-length reverse transcription products are not used as the substrate for integration. If a mutant is defective in catalysis but carries out nuclear import efficiently (such as the D116A mutant), 2LTRc should accumulate with respect to a wt IN. Instead, if the mutant is also impaired in nuclear import, 2LTRc will either not increase with respect to the wt IN or will increase but more modestly than for the D116A mutant.

Hence, to monitor nuclear import, we measured the amount of 2LTRc in wt INA and in mutants containing either no K (NQNQ) or only one (either K<sub>273</sub>, in the NQNK mutant, or K<sub>240</sub>, in the NKNQ mutant). The INA D116A mutant was used as a control. This mutant, being totally inactive for integration, was considered to produce the highest accumulation of 2LTRc, set at 100%. As expected, the level of 2LTRc found with wt INA, which efficiently imports and integrates the reverse-transcribed genome, was significantly lower (25%) than that of the D116A mutant. As shown in Fig. 7A, despite their inability to generate proviral DNA, the mutants had levels of 2LTRc significantly lower than that of the INA D116A mutant, indicative of a defect in nuclear import.

To estimate the efficiency of nuclear import in the mutants, we estimated the level of 2LTRc (Table 1, line 2, "theoretical level"), which could be obtained if no defect in nuclear import was present. We then calculated the efficiency of nuclear import as the ratio between the level of 2LTRc observed experimentally (Table 1, line 3) and the theoretical one. If no defects in nuclear import are present, ratios should be around 1, while defects in nuclear import would yield ratios of <1. The ratios found for the three mutants were in the 0.31 to 0.35 range (Table 1, line 4), indicative of a reduction in nuclear import to approximately one-third that of the wt enzyme. Therefore, the defects in nuclear import contribute to the decrease in integration found with these

**TABLE 1** Estimate of the contribution of the defects of nuclear import and 3' processing to efficiency of integration observed with NKNK motif mutants

Line	Parameter	Value for:				
		wt INA (NKNK)	INA D116A mutant	NQNQ	NQNK	NKNQ
1	Observed levels of integration (relative to wt INA) (values from Fig. 5)	1	0.00 ± 0.00 <sup>a</sup>	0.00 ± 0.00	0.03 ± 0.02	0.24 ± 0.09
2	Theoretical levels of 2LTRc (relative to INA D116A) (see Materials and Methods)	0.20	1	1.00 ± 0.00	0.98 ± 0.02	0.81 ± 0.07
3	Observed levels of 2LTRc (relative to INA D116A) (values from Fig. 7A)	0.20	1	0.33 ± 0.08	0.30 ± 0.09	0.28 ± 0.11
4	Efficiency of nuclear import (relative to INA D116A) (ratios from line 3/values from line 2)	1.00	1	0.33 ± 0.08	0.31 ± 0.09	0.35 ± 0.14
5	Ratio of PJ/2LTRc (relative to INA D116A) (values from Fig. 7B)	0.54 ± 0.02	1	0.93 ± 0.18	0.87 ± 0.19	0.66 ± 0.17
6	Decrease of PJ/2LTRc (relative to INA D116A) (= 1 – values in line 5)	0.46 ± 0.10	0	0.07 ± 0.01	0.13 ± 0.03	0.34 ± 0.09
7	Efficiency of 3' processing (relative to wt INA) (= values in line 6/0.46)	1.00 ± 0.22	0	0.15 ± 0.03	0.28 ± 0.06	0.74 ± 0.19
8	Expected levels of integration (relative to wt INA) (product of values in lines 4 and 7)	1.00 ± 0.22	0	0.05 ± 0.01	0.09 ± 0.03	0.26 ± 0.11

<sup>a</sup>In italics are shown the SDs (calculated as described in Materials and Methods).

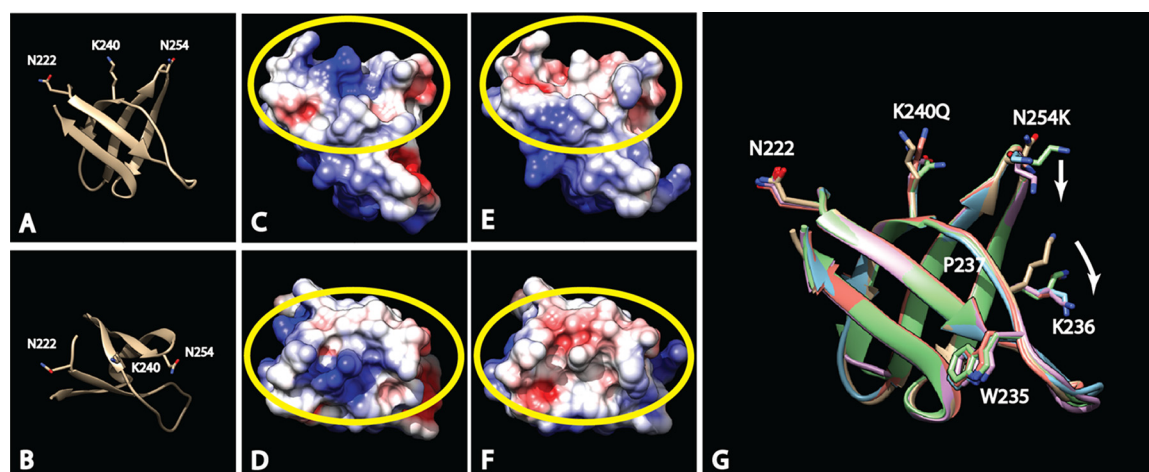
mutants but cannot alone account for the low levels observed, particularly in NQNQ and NQNK mutants, for which integration was almost undetectable (Table 1, line 1).

The efficiency of 3' processing carried out by IN was then analyzed by quantifying the different LTR-LTR junctions in the 2LTRc. The 2LTRc found in the nucleus are generated from DNAs carrying either unprocessed or processed 3' ends. In the first case, the 2LTRc will present "perfect junctions" (PJ), while in the second, the junctions will be "imperfect." A high ratio of PJ/2LTRc is therefore indicative of inefficient 3' processing. We found that mutating both K's (NQNQ mutant) or only K<sub>240</sub> (NQNK mutant) led to results not significantly different from those obtained with the INA D116A catalytic mutant (Fig. 7B), indicative of a marked defect in 3' processing. Mutating K<sub>273</sub> (NKNQ mutant) instead did not affect the process, with PJ/2LTRc values comparable to those of the wt enzyme.

To evaluate the contribution of defects in 3' processing to the decreased integration efficiency observed with the various mutants, we first estimated the maximum diminution of the PJ/2LTRc ratio observed for a fully competent enzyme (wt INA). The PJ/2LTRc ratio for wt INA was 0.54 that of INA D116A (Table 1, line 5), corresponding to a reduction of 46% due to 3' processing (Table 1, line 6). The ratio of PJ/2LTRc with respect to INA D116A was then calculated for each mutant, and the resulting value was divided by 0.46, obtaining an estimate of the efficiency of 3' processing relative to that observed for wt INA (Table 1, line 7). 3' Processing of NQNQ and NQNK mutants was dramatically reduced, to 15% and 28% of that of wt INA, respectively. Mutating K<sub>273</sub>, instead, decreased 3' processing to only 74% of that of wt INA.

Finally, in order to understand if these two types of defects (nuclear import and 3' processing) were sufficient to explain the integration defects observed with the various mutants, we combined the effect of these defects (Table 1, line 8). The values obtained account remarkably well for the efficiencies of integration observed (Table 1, lines 1 and 8), indicating that the decrease observed when mutating the K's of the NKNK motif, once normalized for the differences observed in the amount of viral DNA produced, is essentially due to alterations in these two processes.

**Structural analysis of wt and mutant integrase CTDs.** To understand the structural bases for the functional differences observed in the NKNK motif mutants, the crystal structures of the C-terminal domain (IN CTD, positions 220 to 270) of wild-type INA and of the IN of the reference strain NL4.3 were solved at 2.2 and 1.3 Å of resolution, respectively. For both structures, K<sub>273</sub> was not included, as it is in a disordered region of IN. For all crystal forms, we observed a strong packing interaction through the coordination by the His tag of a nickel ion (data available upon request). The structures had the same topology, consisting of a five-stranded β-barrel (data available upon request). The region encompassing the positions of the motif (Fig. 8A and B) generates a surface endowed with a positive potential (Fig. 8C and D, circled in



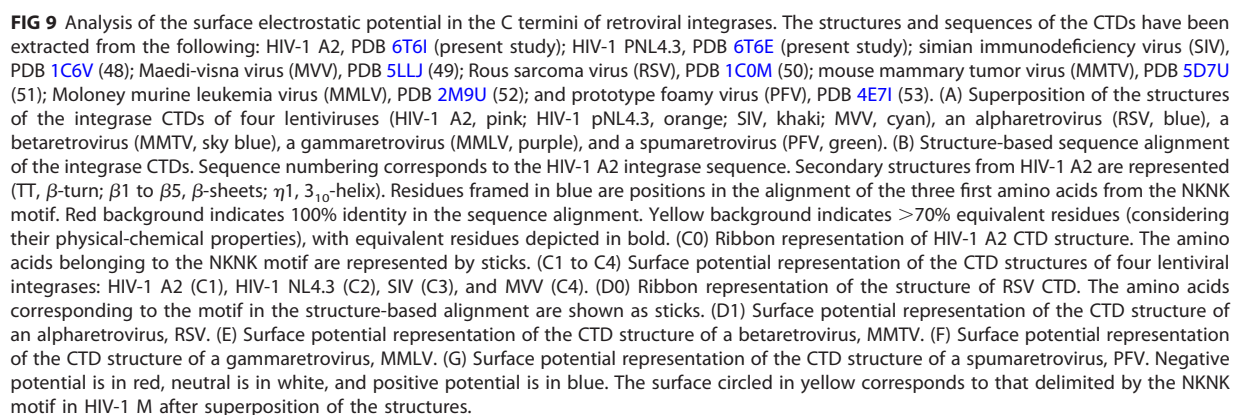
**FIG 8** Structural analysis of the NKNK motif. (A and B) Side view (A) and top view (B) of the ribbon representation of the crystal structure of INA CTD. The positions of residues N222, K240, and N254 are represented by sticks. (C and D) Side view (C) and top view (D) of the surface electrostatic potential representation of INA CTD. (E and F) Side view (E) and top view (F) of the surface electrostatic potential representation of INA NQKK. (G) Superposition of the CTDs of INA and INA NQKK (chains A, B, and C). The mutation N254K induces the displacement of the K236 side chain (white arrows), disturbing the structure of region 235–237. Red, negative potential; blue, positive potential; white, neutral regions. The regions with large differences are circled in yellow.

yellow), suggesting that this feature might be important for the functionality of the IN. In this case, it is expected that inserting additional lysines in the motif (such as with the mutants containing three or four lysines) would retain functionality and that, conversely, removing the K's (mutants with one K or no K) would affect it. This is what we observed, as shown in Fig. 5. Nevertheless, the correlation between surface potential and functionality is less clear for the mutants in which the number of K's in the motif (two) is not altered but their positions are permuted with polar amino acids.

To clarify this point, we solved, at a 2.0-Å resolution, the crystal structure of the CTD of the INA NQKK mutant, which is the one displaying the most dramatic drop in integration among the mutants possessing two K's (Fig. 5C). The INA NQKK CTD crystallized in a different space group and had three chains in the asymmetric unit. The superposition of the five structures corresponding to IN NL4.3 CTD onto INA CTD and onto the three molecules in the asymmetric unit for the INA NQKK CTD (chains A, B, and C) did not show significant differences in the main chain fold (root mean square deviation [RMSD] of atomic positions: IN NL4.3 CTD versus INA CTD = 0.395 Å, INA CTD versus INA NQKK CTD chains A, B, and C = 0.653 Å, INA NQKK CTD chain A versus chain B versus chain C = 0.558 Å). Interestingly, the positive surface electrostatic potential observed for the wt enzyme was markedly perturbed in the INA NQKK mutant (Fig. 8E and F, circled in yellow), a change that might well account for the decrease in functionality of the NQKK mutant integrase.

To further analyze the impact of the mutations on the structure, we defined the regions which are naturally disordered (intrinsically disordered regions [IDRs]) by the superposition of the three molecules in the asymmetric unit of the NQKK CTD structure. We assume that the change in the RMSD obtained among the three molecules represents the natural IDRs. For the main chain, disordered regions with high RMSD are region 228–232 and region 243–248 (data available upon request). These same regions are found to be similarly disordered when the main chain RMSD of the CTD of INA is compared to that of INA NQKK chains A, B, and C (data available upon request), indicating that the mutations have no effect on the C-alpha backbone fold of IN CTD.

The three structures were then analyzed from the standpoint of the arrangement of the side chains. Calculating side chain RMSDs, disordered portions were found to correspond to regions 222–225, 228–232, and 243–248 (data available upon request). The comparison of INA CTD and IN NL4.3 CTD, which differ between positions 220 and 270 by only a single amino acid change (V234 in IN NL4.3 replaced by I234 in INA), as



To evaluate the importance of charge configurations in the context of the CTD of retroviral integrases, we performed an analysis of the electrostatic charge surface potential for other lentiviruses as well as for other retroviruses. Integrase CTD structures are available for the following: HIV-1 A2, PDB [6T6I](#) (present study); HIV-1 PNL4.3, PDB [6T6E](#) (present study); simian immunodeficiency virus (SIV), PDB [1C6V](#) (48); Maedi-visna virus (MVV), PDB [5LJJ](#) (49); Rous sarcoma virus (RSV), PDB [1C0M](#) (50); mouse mammary tumor virus (MMTV), PDB [5D7U](#) (51); Moloney murine leukemia virus (MMLV), PDB [2M9U](#) (52); and prototype foamy virus (PFV), PDB [4E7I](#) (53). First, we superposed the available structures. The superposition shows that they share a fold (Fig. 9A) as well as a low RMSD on the secondary-structure backbone despite levels of sequence identity for some cases as low as 10% (data available upon request). A structure-based sequence alignment was then performed (Fig. 9B). Surprisingly, despite a low overall sequence



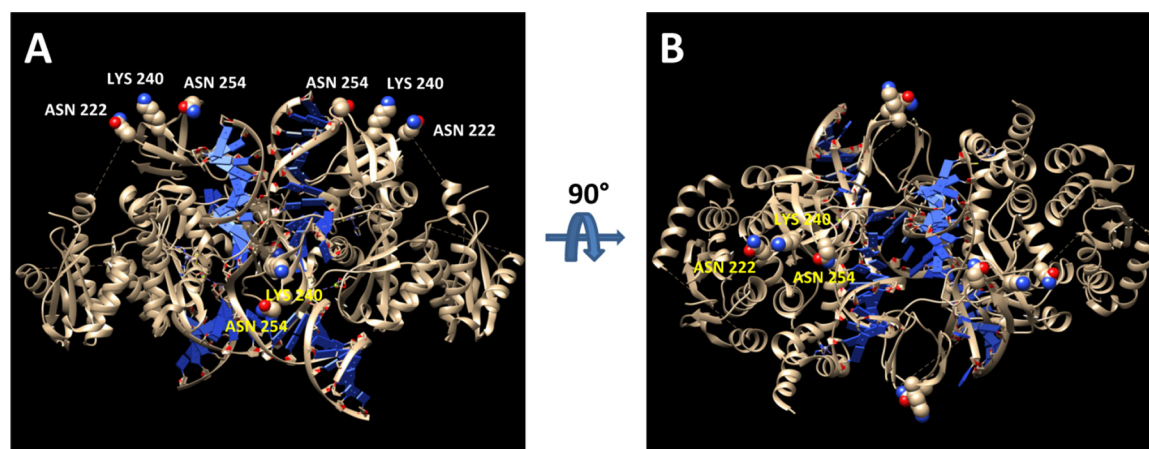
identity (10% to 20%) for some integrases, several regions have a strong local sequence similarity (Fig. 9B, red and yellow highlighting), while no conservation is observed at positions 222, 240, and 254 among lentiviruses or retroviruses. However, when we compared the electrostatic surface potentials of all structures, we could define two general retroviral classes: a first class represented by lentiviruses (Fig. 9C1 to C4) in which the surface corresponding to the one delimited by the NKNK motif in HIV-1 M is basic, and a second class represented by the other retroviruses tested (alpha-, beta-, and gammaretroviruses and spumaretroviruses; Fig. 9D1 to G) in which this surface is acidic or neutral.

## DISCUSSION

Here, by performing a systematic comparison between the nonconserved amino acids in the CTDs of the HIV-1 group M and group O integrases, we identified, in group M, a highly conserved motif that is essential for integration. The motif is constituted by two asparagines and two lysines (N<sub>222</sub>K<sub>240</sub>N<sub>254</sub>K<sub>273</sub>), all required for the generation of proviral DNA. In particular, when the K's were mutated, integration was abolished due to the cumulative effects of decreased 3' processing and nuclear import of the reverse transcription products (Table 1). Replacing the K with R did not affect integration (Fig. 3E), suggesting that the essential feature of the K's is their positive charge. Importantly, the positions of the two K's of the motif could be permuted without affecting the functionality of the integrase in most cases (Fig. 5).

A potential explanation for the retention of functionality when permutating the positions of the K's across the motif comes from the structural data of the CTD obtained in this work. We have solved the crystal structure of the CTD of the wt INA used in this study as well as that of the K240Q N254K mutant (referred to as NQKK in Results). In the structure of the wt enzyme, the residues constituting the motif (except K<sub>273</sub>, which is part of an unresolved region) generate a positively charged surface (Fig. 8A to D). This positive electrostatic potential surface is absent, in contrast, in the NQKK mutant (Fig. 8E and F), which despite the presence of two K's, displays a drastic reduction in integration efficiency. These results, combined with the tests of functionality of the different mutants, suggest that the relevant parameter is the presence of a positive charge across this surface. Charged residues have a strong effect on the surface potential. The nature of the amino acid side chain (charged, polar, nonpolar) on the surface of the protein defines the surface potential. Charged and polar groups, through formation of ion pairs, hydrogen bonds, and other electrostatic interactions, impart important properties to proteins. Modulation of the charges on the amino acids, e.g., by pH and posttranslational modifications, have significant effects on protein-protein and protein-nucleic acid interactions (54). In addition to residues carrying net charges, polar residues also have significant partial charges and can form hydrogen bonds and other specific electrostatic interactions among themselves and with charged residues (55). In the case of the present study, the possible contribution of these mechanisms to the functionality of the integrase might be reflected by the loss of functionality observed by replacing the N, which carries an amide side chain (–CONH<sub>2</sub>), with either a nonpolar amino acid (L) or a polar one (T) but carrying a hydroxyl side chain (–OH). The analysis of the electrostatic surface potential for the integrase CTD of the retroviruses for which this is known showed that the topology of the structure is maintained (Fig. 9A), and the analysis of the surface electrostatic potential splits the viruses studied into two classes. One consists of lentiviruses, for which the surface delimited by the NKNK motif of HIV-1 M contains basic charges (in some cases brought by amino acids noncorresponding to those of the NKNK motif of HIV-1 M). The second class includes the other orthoretroviruses studied (alpha-, beta-, and gammaretroviruses) and spumaretroviruses, for which this surface contains acidic and neutral regions. This presence of basic regions, specifically in lentiviruses, might contribute to some specificity of lentiviral biology to increase the efficiency of infection of quiescent cells.

The importance for protein functionality of charge configurations and clusters in their three-dimensional structures has been underlined by several studies (56–59). Charge permutations have been used in the NC region of the Gag protein for the



**FIG 10** The NKNK motif in the postintegration intasome structure. Two 90° views of the intasome structure from Passos et al. (31) are shown. The residues forming the electrostatic surface are pointing toward the solvent for the two independent CTDs of the integrase dimer.

Mason-Pfizer monkey virus (60). This basic region could be replaced with nonspecific sequences containing basic amino acid residues, without altering its functionality, while mutants with neutral or negatively charged residues showed a large drop in viral infectivity in single-round experiments. Moreover, a mutant exhibiting an increased net charge of the basic region was 30% more infectious than the wild type. Also, in our study, increasing the positive charge of the NKNK motif of HIV-1 IN by introducing a third K leads to a slight increase in integration with respect to wt IN (Fig. 5A and D).

As retention of IN functionality relies on the electrostatic surface potential rather than on the specific positions of the positively charged amino acid, we infer that this region is probably involved in an interaction with a partner carrying a repetitive negatively charged biochemical motif, such as the phosphates of the nucleic acid backbone. Alternatively, the partner could be a disordered region of a protein that can rearrange to preserve the interaction when the positions of the positive charges are permuted across the surface of the NKNK motif. Indeed, the molecular recognition between charged surfaces and flexible macromolecules like DNA, RNA, and intrinsically disordered protein regions has been observed for prototype foamy virus and Rous sarcoma virus Gag precursors (61, 62), for EBNA proteins of Epstein-Barr virus (63), and the UL34 protein of herpes simplex virus (64), as well as for cellular proteins like APOBEC3G (65). Moreover, the presence of asparagines in the motif, which we show are required for integrase functionality, might contribute to the interaction with the nucleic acid or with a protein partner through hydrogen bonds with the bases (54) or with polar amino acids (55, 66, 67). The analysis of the motif in the context of the well-characterized structure of the intasome (31), mimicking a postintegration deoxyribo-nucleic complex, indicates that the residues forming the electrostatic surface point toward the solvent, at the exterior of the structure (Fig. 10). This is coherent with the observation that mutating the motif does not affect late steps of the integration process but rather earlier ones such as 3' processing and nuclear import.

We show that the NKNK motif of the CTD is involved in 3' processing and nuclear import of the reverse transcription product. Indeed, the removal of the K impacts both processes, and when combined, these effects are sufficient to account for the drop of infectivity to the undetectable levels observed in the absence of K in the motif (Table 1). Since mutating K<sub>240</sub> has a strong impact on both 3' processing and nuclear import, while K<sub>273</sub> appears to be predominantly involved in nuclear import, it is possible that the involvement of the motif in these two processes implicates structurally distinct functional complexes. This is the first finding of an implication of the HIV-1 IN CTD in 3' processing. So far, only the involvement of the catalytic domain had been demonstrated (39, 68, 69). Since it has been shown that different oligomerization states



of IN specifically influence the ability to carry out 3' processing or strand transfer (70), it is possible that the electrostatic surface formed by the NKNK motif helps stabilize the oligomeric state that allows 3' processing.

Concerning nuclear import, it is known that HIV-1 IN binds several cellular nuclear import factors through basic amino acids of the CTD and that abolishing these interactions leads to noninfectious viruses displaying a severe defect in nuclear import. Here, we extend the regions of IN involved in this process by describing the need for a new motif, although it cannot be determined whether its involvement is direct or mediated by the interaction with a partner protein with karyophilic properties.

Some of the residues constituting the  $N_{222}K_{240}N_{254}K_{273}$  motif have been previously characterized as showing their implication in different steps of the infectious cycle. One is the involvement in reverse transcription. The integrase CTD interacts with the reverse transcriptase to improve DNA synthesis (71, 72). In one study, the double mutation K240A K244E caused a decrease in reverse transcription to around 20% of the levels of the wt enzyme (72), while the K244E mutation alone caused a reduction of 40% in RT efficiency (73), suggesting that  $K_{240}$  also contributes to reverse transcription. The decrease we observed when mutating  $K_{240}$  alone, to around 45% of wt activity, is consistent with this view. The characterization by nuclear magnetic resonance (NMR) of the RT-binding surface in the IN CTD, obtained using the CTD comprising positions 220 to 270, shows that it is made up of 9 residues (amino acids 231 to 258, including  $K_{244}$ ) that strongly interact both with the RT alone (9) and with the RT-DNA complex (74). When the interaction involves the complex, this surface includes 5 additional amino acids (74). Among these additional residues are  $N_{222}$  and  $K_{240}$ , which are located at one edge of the surface. It is therefore possible that the nature of the residues at positions 222 and 240 affects the interaction between the CTD and RT-DNA complex.

Concerning  $K_{273}$ , contradictory results have been obtained for reverse transcription of viruses harboring integrases with sequential C-terminal deletions ( $IN_{1-270}$  and  $IN_{1-272}$ ) (75, 76). Furthermore, for reverse transcription to occur, the genomic RNA must be encapsidated in the core of the viral particle. With respect to this, it has been recently shown that mutating  $K_{273}$  together with  $R_{269}$  (R269A K273A mutant) impairs encapsidation of the genomic RNA (6). As expected, reverse transcription in the double mutant was almost abolished. Here, mutating  $K_{273}$  to Q led to a reduction in reverse transcription of only 30% (Fig. 3B), suggesting that mutating  $K_{273}$  alone is not sufficient to affect genomic incorporation into the viral capsid, at least in the majority of particles. Supporting this, an earlier study (77) showed that the K273A single mutation did not affect viral replication in Jurkat cells, indicating that it is the specific combination of R269A and K273A mutations that is responsible for the impairment of the genome encapsidation.

Finally, acetylation of  $K_{273}$  has been previously proposed to be important for different steps of the infectious cycle (78, 79). In those studies, though, the role of acetylation of  $K_{273}$  was assessed by simultaneously replacing  $K_{264}$ ,  $K_{266}$ , and  $K_{273}$ , thereby not allowing conclusions regarding the specific contribution of  $K_{273}$ . Here, hampering acetylation but preserving the positive charge by the K273R substitution did not affect integration, indicating that the possible acetylation of  $K_{273}$  had no effect on integration. This observation is in line with the observation by Topper and coworkers that posttranslational acetylation of the integrase CTD is dispensable for viral replication (80). All together, the data available in the literature regarding  $K_{240}$  and  $K_{273}$  indicate that the effects we observed in this study cannot be due to any of the already known properties of the residues of the motif.

The NKNK motif is strictly conserved in natural sequences of HIV-1 group M. However, we show that various variants of the NKNK motif display levels of integration efficiency equivalent to the wt enzyme and could therefore in principle be found in the epidemics. Their absence is indicative of purifying selection occurring *in vivo*, likely exerted at a step different from integration. One possibility is the implication of IN in reverse transcription, which is, in all variants, less efficient than with the NKNK sequence. The existence of several alternative sequence arrangements in the motif with

comparable efficiencies of integration might have therefore constituted an asset for optimizing the acquisition of additional functions, such as promoting reverse transcription.

## MATERIALS AND METHODS

**Plasmids and molecular cloning.** p8.91-MB was constructed by engineering one *Mlu*I and one *Bsp*EI restriction site 18 nt downstream of the 5' end and 21 nt upstream of the 3' end, respectively, of the RT coding sequence of pCMVΔR8.91 (81). The insertion of the two restriction sites led to three amino acid changes in the RT (E6T, T7R, and A554S). These modifications only slightly affected the efficiency of generation of puromycin-resistant clones (see below) upon transduction with the resulting viral vector (v8.91-MB), since in three independent experiments, the number of clones obtained with p8.91-MB was  $80\% \pm 6\%$  of that obtained using p8.91. p8.91-MB was employed throughout the study as a positive control and was used to insert the various variants of RT and IN tested. Together with the *Sall* site, present in p8.91 48 bp downstream of the stop codon of the *pol* gene coding DNA sequence (CDS), the *Mlu*I and *Bsp*EI sites define two exchangeable cassettes: one encompassing the RT coding sequence (*Mlu*I-RT-*Bsp*EI, 1,680 bp) and one encompassing the IN-coding sequence (*Bsp*EI-IN-*Sall*, 940 bp). These cassettes were used to insert the various sequences of RT and IN used in the study. The plasmid used to produce the genomic RNA of the viral vectors was a modified version of pSDY, previously described (82), here called pSRP (for pSDY-nRFP-Puro). This variant was obtained by introducing two modifications into the original pSDY-dCK-Puro plasmid (82). The first one was the replacement of the sequence encoding human deoxycytidine kinase with a cassette containing the red fluorescent protein (RFP) fused with the N-terminal 124 amino acids of human histone H2B, which directs the RFP to the nucleus. The RFP was used to monitor the efficiency of transfection by fluorescence microscopy. The second modification was the replacement of the HIV-1 U3 sequence in the 5' LTR with that of the U3 of the Rous sarcoma virus. For the generation of qPCR standard curves, two plasmids were constructed: one, called pJet-1LTR, for the detection of early and late reverse transcription products, was obtained by inserting the sequence encompassing the LTR and the Psi region from pSDY (82) into the pJET plasmid with the CloneJET PCR cloning kit (Thermo Scientific, MA, USA); the second, pGenuine2LTR, was obtained by inserting a fragment of 290 bp corresponding to the unprocessed junction of U5/U3 (CAGT/ACTG being the sequence of the junction 5' to 3') into the pEX-A2 plasmid (Eurofins Genomics, Luxembourg). For the study with replication-competent viruses, we used the pNL4.3 plasmid (83), which was obtained from the NIH AIDS Research and Reference Reagent Program, no. 114 (GenBank accession no. [AF324493](#)). We replaced in this plasmid the coding sequence of NL4.3 IN CTD with those of wt and mutant INA CTD, as described in Results. Chimeric integrases between primary isolates from HIV-1 group M subtype A2 and HIV-1 group O RBF206, as well as mutant integrases, were constructed through overlap extension PCR as previously described for the envelope gene (84).

**Cells.** HEK-293T cells were obtained from the American Type Culture Collection (ATCC). P4-CCR5 reporter cells are HeLa CD4<sup>+</sup> CXCR4<sup>+</sup> CCR5<sup>+</sup> carrying the LacZ gene under the control of the HIV-1 LTR promoter (85). TZM-bl cells are an HeLa cell clone genetically engineered to express CD4, CXCR4, and CCR5 and containing the Tat-responsive reporter gene for firefly luciferase under the control of the HIV-1 LTR (86). HEK-293T, P4-CCR5, and TZM-bl cells were grown in Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher, MA, USA) supplemented with 10% fetal calf serum and 100 U/ml penicillin–100 mg/ml streptomycin (Thermo Fisher, MA, USA) at 37°C in 5% CO<sub>2</sub>. CEM-SS cells are human T4-lymphoblastoid cells (87–89) and were grown in RPMI medium supplemented with 10% fetal calf serum and 100 U/ml penicillin–100 mg/ml streptomycin (Thermo Fisher, MA, USA) at 37°C in 5% CO<sub>2</sub>.

**Viral strains.** The following primary isolates were used for this study: from HIV-1 group M, one from subtype A2 (GenBank accession no. [AF286237](#), here named isolate A), one from subtype C (GenBank accession no. [AF286224](#), here named isolate C), one from CRF02\_AG (GenBank accession no. [MH351678](#)), and one from subtype B (isolate AiHo; GenBank accession no. [MH351679](#), here named isolate B); from HIV-1 group O, the primary isolate RBF206 (GenBank accession no. [KU168298](#), here named isolate O). Isolates in GenBank under accession no. [AF286237](#), [AF286224](#) and [MH351678](#) were obtained from the NIH AIDS Research and Reference Reagent Program; isolates in GenBank under accession no. [MH351678](#), [MH351679](#) and [KU168298](#) were kindly provided by J. C. Plantier (CHU Rouen, France).

**Sequence alignments.** We used 3,366 HIV-1 sequences for alignment. HIV-1 group M sequences were downloaded from the Los Alamos National Laboratory (LANL) HIV sequence database and correspond to the different HIV-1 group M pure subtypes as follows: A (249 sequences), B (2,450 sequences), C (450 sequences), D (121 sequences), G (80 sequences), H (8 sequences), J (6 sequences), and K (2 sequences). We also aligned 49 HIV-1 group O sequences, using 26 sequences from the LANL database and 23 sequences obtained through collaboration with the Virology Unit associated with the French National HIV Reference Center (J. C. Plantier, CHU Rouen). Sequence alignments were performed with CLC sequence viewer 8. The sequence logo of positions 222, 240, 254, and 273 in HIV-1 group M IN was obtained with an alignment of 3,366 sequences of the IN CTD using WebLogo (<http://weblogo.berkeley.edu/logo.cgi>).

**Generation of pseudotyped viral vectors.** Pseudotyped lentiviral vectors were produced by cotransfection of HEK 293T cells with pHCMV-G (90) encoding the VSV-G envelope protein, pSRP, and p8.91-MB-based plasmids by the polyethylenimine method in accordance with the manufacturer's instructions (PEI, MW 25000, linear; Polysciences, Warrington, PA, USA). HEK 293T cells were seeded at  $5 \times 10^6$  per 100-mm diameter dish and transfected 16 to 20 h later. The medium was replaced 6 h after

**TABLE 2** Primers and probes used in the qPCR assay

Aim	Target	Primer/probe or description	Sequence (5'–3')	Fluorophore <sup>a</sup>
Quantification	U5Psi	U5Psi-forward	GTGACTCTGGTAACTAGAGA	FAM
	U5Psi	U5Psi-probe	CGCTTCAAGTCCCTGTTCCGG	
	U5Psi	U5Psi-reverse	GAGAGCTCCTCTCCTTC	
	RU5	RU5-forward	CAGATCTGAGCCTGGGAG	HEX
	RU5	RU5-probe	AAGCAGTGGGTCCCTAGTTAGCC	
	RU5	RU5-reverse	GGCACACACTACTTGAAGC	
Normalization	ACTB	IDT predesigned assay, Hs.PT.56a.40703009.g/exon 6		HEX

<sup>a</sup>FAM, 6-carboxyfluorescein; HEX, 6-carboxy-2,4,4,5,7,7-hexachlorofluorescein.

transfection, the vectors were recovered from the supernatant 72 h later and filtered on 0.45- $\mu$ m filters, and the amount of p24 (CA) was quantified by ELISA (Fujirebo Europe, Belgium).

**Western blot analysis.** Western blot analysis was carried out on virions to assess the proteolytic processing of the Pr55Gag polyprotein. A 1.5-ml volume of viral supernatant was centrifuged through 20% sucrose, and the virion pellet was lysed in 1.5 $\times$  Laemmli buffer. Viral proteins were separated on a Criterion TGX Stain-Free 4-to-15% gradient gel (Bio-Rad, CA, USA) (0.025 M Tris base, 0.192 M glycine, 0.1% SDS [TGS]; 150 V, 45 min), blotted on a polyvinylidene difluoride (PVDF) membrane (TGS-10% ethanol, 200 mA, 1.5 h), and probed with a mouse monoclonal anti-CA antibody (NIH AIDS Reagent Program, no. 3537) to detect the viral capsid, the Pr55Gag unprocessed polyprotein, and CA-containing proteolytic intermediates. An anti-mouse horseradish peroxidase (HRP)-conjugated secondary antibody was used to probe the membrane previously incubated with anti-CA antibody. Membranes were incubated with ECL reagent (Thermo Fisher, MA, USA), and Western blots were imaged on a Bio-Rad ChemiDoc Touch system and analyzed with Bio-Rad Image Lab software.

**Evaluation of reverse transcription by qPCR.** The viral vectors were treated with 200 U/ml of Benzonase nuclease (Sigma-Aldrich, MO, USA) in the presence of 1 mM MgCl<sub>2</sub> for 1 h at 37°C to remove noninternalized DNA. The vectors (200 ng of p24) were then used to transduce  $0.5 \times 10^6$  HEK 293T cells by spinoculation for 2 h at 32°C, 800 relative centrifugal force (rcf), with 8  $\mu$ g/ml Polybrene (Sigma-Aldrich, MO, USA). After 2 h, the supernatant was removed and cells were resuspended in 2 ml of DMEM and plated in 6-well plates. After 30 h, cells were trypsinized and pelleted. Total DNA was extracted using an UltraClean tissue and cell DNA isolation kit (Ozyme, France). A duplex qPCR assay (primers shown in Table 2) was used to quantify early and late reverse transcription products by detecting the R-U5 and U5-Psi junctions, respectively, and another qPCR (Table 2) was used to normalize the quantity of cells employed in the assay (detection of  $\beta$ -actin exon 6 genomic DNA; International DNA Technologies, Belgium). All primers and probes were synthesized by IDT. The qPCR assays were designed with the TaqMan hydrolysis probe technology using the IDT Primers and Probes design software, with dual quencher probes (one internal ZEN quencher and one 3' Iowa Black FQ quencher) (Table 2). qPCRs were performed with the iTaq universal probe supermix (Bio-Rad, CA, USA) on a CFX96 (Bio-Rad, CA, USA) thermal cycler with the following cycling conditions: initial Taq activation for 3 min at 95°C, followed by 40 cycles of denaturation for 10 s at 95°C and elongation for 20 s at 55°C. Standard curves and analysis were carried out with the CFX Manager (Bio-Rad, CA, USA). DNA copy number was determined using a standard curve prepared with serial dilutions of the reference plasmid pJet-1LTR and of a known number of HEK 293T cells.

**Evaluation of integration by puromycin assay.** Half a million HEK 293T cells were transduced with a volume of viral vectors corresponding to 0.2 ng of p24, by spinoculation for 2 h at 32°C, 800 rcf, with 8  $\mu$ g/ml Polybrene (Sigma-Aldrich, MO, USA). After 2 h, the supernatant was removed and cells were resuspended in 7 ml of DMEM and plated in 100-mm-diameter plates. After 30 h, puromycin was added at a final concentration of 0.6  $\mu$ g/ml, and clones were allowed to grow for 10 to 12 days and then counted. However, the number of clones depends on two parameters: the efficiency of integration and the amount of proviral DNAs available for integration (which depends on the efficiency of reverse transcription). Therefore, we normalized the number of clones observed by the amount of viral DNA generated by reverse transcription (estimated by qPCR) to extrapolate the efficiency of integration. The percentage of integration efficiency for sample X with respect to the control C is, thus, given by  $[(p_x/r_x)/(p_c/r_c)] \times 100$ , where  $r_x$  and  $r_c$  are the amounts of late reverse transcription products, estimated by qPCR, in sample X and control C, respectively, and  $p_x$  and  $p_c$  are the numbers of puromycin-resistant clones in sample X and control C, respectively.

**Evaluation of integration by Alu qPCR.** Equal amounts of total DNA extracted from transduced cells (as deduced by qPCR of  $\beta$ -actin exon 6 genomic DNA; see above) were used for the Alu PCR assay, as previously described (91). Two subsequent amplifications were performed. The first one, comprising 95°C for 3 min, followed by 15 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 3 min 30 s, and then 72°C for 7 min, using the Alu forward primer and the Psi reverse primer, allowed the amplification and quantification of Alu LTR fragments (Table 3). Samples were then diluted to 1:10, and 2  $\mu$ l was used for the second amplification to detect the viral LTR, as described above for the detection of the R-U5 junction. The percentage of integration efficiency for sample X with respect to the control C is given by  $[(a_x/r_x)/(a_c/r_c)] \times 100$ , where  $r_x$  and  $r_c$  are the amounts of late reverse transcription products, estimated by qPCR, in sample X and control C, respectively, and  $a_x$  and  $a_c$  are the amounts of DNA estimated by the second amplification of the Alu qPCR assay in sample X and control C, respectively.

**Quantification of two LTR circles and of circles with perfect junctions.** Noninternalized DNA was removed by treatment with Benzonase nuclease as described for the qPCR assay, and  $0.5 \times 10^6$  HEK 293T

**TABLE 3** Primers used for the Alu qPCR assay

Stage	Target	Primer/probe	Sequence (5'–3')	Fluorophore
1st PCR	Alu LTR	Alu forward	TGCTGGGATTACAGGCGTGAG	
	Alu LTR	Psi reverse	GCTCCTCTGGTTCCCTTTC	
2nd qPCR	RU5	RU5-forward	CAGATCTGAGCCTGGGAG	HEX
	RU5	RU5-probe	AAGCAGTGGGTTCCCTAGTTAGCC	
	RU5	RU5-reverse	GGCACACACTACTTGAAGC	

cells were transduced with a volume of viral vectors corresponding to 1  $\mu$ g of p24 by spinoculation, as described above. After 30 h, cells were trypsinized and pelleted. Total DNA was extracted with an UltraClean tissue and cell DNA isolation kit. Late reverse transcription products (detection of the U5-Psi junction) were quantified as described above, and two qPCR assays were used to quantify the 2LTRc and the quantity of 2LTR circles with a perfect palindromic junction, with a primer overlapping the 2LTRc junction, as previously described (92). The qPCR assays were designed and the primers and probes (Table 4) were synthesized as described above. qPCRs were performed as described above. Standard curves and analysis were carried out with the CFX Manager (Bio-Rad, CA, USA). Copy numbers of the different forms of viral DNA were determined with respect to a standard curve prepared by serial dilutions of the pGenuine2LTR plasmid. The amount of 2LTRc for sample X is normalized by the total amount of viral DNA (detection of the late reverse transcription products by quantifying the U5-Psi junction), and then it is expressed as a percentage of the amount detected for the control INA D116A mutant (represented by D), thus giving  $[(2LTR_{cX}/r_X)/(2LTR_{cD}/r_D)] \times 100$ , where  $r_X$  and  $r_D$  are the amounts of late reverse transcription products in sample X and control D, respectively, and  $2LTR_{cX}$  and  $2LTR_{cD}$  are the amounts of 2LTRc in sample X and control D, respectively.

**Calculation of the efficiency of nuclear import and of 3' processing.** The efficiency of nuclear import was estimated as follows. The level of 2LTRc found with wt INA was 0.2 with respect to that found with the D116A mutant (data from Fig. 7A). The diminution observed with wt INA with respect to the D116A mutant (which was considered to produce the maximum amount of 2LTRc and was therefore set at 1) was therefore 0.8 (given by  $1 - 0.2$ ). Since the diminution of 2LTRc is proportional to the efficiency of integration, for example, a mutant integrating with an efficiency of 0.3 of that of wt INA is expected to reduce the amount of 2LTRc by  $0.8 \times 0.3 = 0.24$ . The amount of 2LTRc expected for that mutant would therefore be given by  $1 - 0.24 = 0.76$ . Similarly, a mutant integrating with a higher efficiency (for example, 0.9 of that of wt INA) is expected to give  $1 - (0.8 \times 0.9) = 0.28$  2LTRc with respect to the D116A mutant. Therefore, the formula applied to estimate the expected levels of 2LTRc with respect to the D116A mutant for sample  $n$  is given by  $1 - (0.8 \times a_n)$ , where  $a_n$  is the level of integration measured for sample  $n$ , relative to wt INA (data from Fig. 7A). The values of 2LTRc measured experimentally (Table 1, line 3) are then divided by the expected ones to obtain an estimate of the relative efficiency of nuclear import (Table 1, line 4).

The efficiency of 3' processing was calculated as follows. The ratio of perfect junctions to the total amount of 2LTRc (PJ/2LTRc) found for the D116A mutant was considered to be the maximal one and was therefore assigned a value of 1 (Table 1, line 5). The proportion of PJ/2LTRc found for wt INA (Table 1) was 0.54 that of the D116A mutant (Table 1, line 5). The proportion by which the pool of PJ/2LTRc found with a catalytically inactive IN can be decreased by 3' processing carried out by a fully catalytic active IN is therefore  $1 - 0.54 = 0.46$  (Table 1, line 6). For the NQNK mutant, for example, the PJ/2LTRc ratio observed was 0.87 of that of the D116A mutant, which corresponds to a relative decrease of the PJ/2LTRc pool of 0.13 (Table 1, line 6). This decrease is  $0.13/0.46 = 0.28$  of that observed for wt INA (Table 1, line 7), providing an estimate of the relative efficiency of 3' processing by this mutant with respect to wt INA. The general formula we applied to estimate the efficiency of 3' processing was therefore  $(1 - r_X)/0.46$ , where 1 is the proportion of PJ/2LTRc found for the D116A mutant,  $r_X$  is the PJ/2LTRc ratio observed for sample X, and 0.46 is the decrease in the PJ/2LTRc ratio observed for wt INA with respect to the D116A mutant. The resulting values are reported in Table 1, line 7.

In Table 1, the standard deviations (SDs) are given in italics. In lines 1, 3, and 5, SDs are derived from the experimental values; in line 2, SD is given by  $1 - 0.2$  multiplied by the corresponding SD from line 1; in line 4, SD is given by  $[(SD_{line3}/average_{line3})^2 + (SD_{line2}/average_{line2})^2]^{1/2} \times average_{line4}$ ; in line 6, SD

**TABLE 4** Primers and probes used for the detection of total 2LTRc and 2LTRc with perfect palindromic junction

Target <sup>a</sup>	Primer/probe	Sequence (5'–3')	Fluorophore
2LTRc	2LTR forward	CCCTTTAGTCAGTGTGGAA	FAM
	2LTR probe	TTCACTCCCAACGAAGACAAGATATCCTT	
	2LTR reverse	GTAGCCTTGTGTGTGGTAGA	
PJ	2LTR PJ forward	TGTGGAAAAATCTCTAGCAGTAC	FAM
	2LTR probe	TTCACTCCCAACGAAGACAAGATATCCTT	
	2LTR reverse	GTAGCCTTGTGTGTGGTAGA	

<sup>a</sup>PJ, perfect junction.

is given by  $(SD_{line5}/average_{line5}) \times average_{line6}$ ; in line 7, SD is given by  $SD_{line6}/0.46$ ; in line 8, SD is given by  $[(SD_{line7}/average_{line7})^2 + (SD_{line4}/average_{line4})^2]^{1/2} \times average_{line8}$ .

**Assessment of infectivity of replication-competent viruses.** As described above, the coding sequence of NL4.3 IN CTD was replaced with those of wt and mutant INA CTDs. Replication-competent viruses were produced as described above, and equal amounts of viruses were used to infect cells (TZM-bl or CEM-SS) for each sample. For estimating viral replication in TZM-bl cells, 25  $\mu$ l of virus dilution was added to  $10^4$  cells, plated in 96-well plates in 75  $\mu$ l of culture medium. After 48 h, virus replication was detected by measuring Luc reporter gene expression by removing 50  $\mu$ l of culture medium from each well and adding 50  $\mu$ l of Bright Glo reagent to the cells. After 2 min of incubation at room temperature to allow cell lysis, 100  $\mu$ l of cell lysate was transferred to 96-well black solid plates for measurements of luminescence (relative light units [RLU]) using a luminometer (93). For the detection of virus replication in CEM-SS cells,  $0.5 \times 10^6$  CEM-SS cells/5 ml were infected with 1/25 virus dilution. After 5 days of culture, the percentage of infected cells was detected by intracellular p24 immunostaining and flow cytometry analysis as previously described (94).

**Cloning, production, purification, and crystallization.** The C-terminal domains (CTDs; positions 220 to 270) of IN NL4.3, INA, and INA NQKK studied here were cloned into the pET15b plasmid, and the proteins were expressed in *E. coli* BL21(DE3) cells. After transformation with the IN C-terminally expressing pET15b, bacteria were inoculated at an optical density at 600 nm ( $OD_{600}$ ) of 0.1 in 1 liter of LB medium supplemented with 10% (wt/vol) sucrose. Cultures were incubated at 37°C with shaking at 220 rpm. At an  $OD_{600}$  of 0.5, the temperature was lowered to 25°C, and shaking was reduced to 190 rpm, until the cells reached an  $OD_{600}$  of 0.8. IPTG (isopropyl- $\beta$ -D-thiogalactopyranoside) was then added to a final concentration of 0.5 mM to induce the expression of the CTDs. Cells were incubated overnight at 25°C. Bacteria were then collected by centrifugation.

For protein purification, cells were resuspended in lysis buffer (25 mM HEPES, pH 8, 1 M NaCl, 10 mM imidazole) at a ratio of 10 ml of buffer/g of biomass. Roche cOmplete inhibitor cocktail tablets were added at the beginning of lysis to avoid protease degradation. Cells were lysed by sonication for 1 min/g of cells with a pulse every 2 s at 40% amplitude at 4°C. The bacterial debris was pelleted by ultracentrifugation at  $100,000 \times g$  for 1 h at 4°C. The supernatant was then loaded onto a 1-ml HisTrap FF Crude column (GE Healthcare) at a flow rate of 1 ml/min using the AKTA fast protein liquid chromatograph (FPLC). Protein was eluted using a gradient of up to 500 mM imidazole in 10 column volumes. Protein concentration was estimated using the NanoDrop spectrophotometer. Subsequently, the protein sample was concentrated using an Amicon Ultra 15 ml with a 3-kDa molecular weight cutoff (MWCO) for the next purification step. A second step of purification was carried out using a S75-16/60 column (GE Healthcare) in 25 mM HEPES, pH 8, 1 M NaCl. Samples were dialyzed into 25 mM HEPES, pH 8, 150 mM NaCl for crystallization.

All initial crystallization conditions were determined by vapor diffusion using the TPP Labtech Mosquito Crystal. Two hundred nanoliters of protein (7–4 mg/ml) was mixed with 200 nl of reservoir contents (described below) in two or three wells of a 96-well MRC crystallization plate, which was stored in a Formulatrix RockImager at 20°C. Screens included PEGs (Hampton Research), MPD, CLASSICS, NUCLEIX (Qiagen), JCSG, WIZARDS, ANION, and CATION (Molecular Dimensions). Once the initial conditions were obtained, manual drops were set up in a Hampton Research 24-well VDX plate to optimize crystallization conditions and to improve crystal size and quality by mixing 1  $\mu$ l protein plus 1  $\mu$ l reservoir contents and equilibration against 500  $\mu$ l of reservoir contents at 20°C. IN CTD NL4.3 (group M, subtype B) was obtained in a reservoir containing 0.1 M Tris, pH 7, 0.8 M potassium sodium tartrate, and 0.2 M lithium sulfate. For INA CTD (subtype A2) and INA NQKK CTD, the reservoir contained 0.1 M MES, pH 6.5, and 1 M sodium malonate.

**Data collection, structure solving, and refinement.** Data collection was performed at the Swiss Light Source (SLS, Villigen, Switzerland) on a Dectris Pilatus 2M detector. After fishing, crystals were rapidly passed through a drop of fluorinated oil (Fomblin Y LVAC 14/6; average MW of 2,500 [Sigma-Aldrich]) to prevent ice formation and directly frozen on the beamline in the nitrogen stream at 100 K. X-ray diffraction images were indexed and scaled with XDS (95, 96). The structures were solved by molecular replacement using PHASER (97) in the PHENIX (98) program suite using the NMR HIV-1 C-terminal structure (PDB 1QMC) (99) as a search model for the IN CTD NL4.3 structure, which was used subsequently as a search model to solve the INA and INA NQKK structures. The structure was then built using the AUTOBUILD program (100, 101), followed by several cycles of refinement using PHENIX.REFINE (102) and manual rebuilding with COOT (103). Structure-based sequence alignment was performed using PROMALS3D (104). Structure superposition and root mean square deviation (RMSD) calculations were performed using secondary structure matching (SSM), superpose program (105) embedded in COOT (103) and in the CCP4 program suite (106). The sequence alignment representation was generated by ESPript (107). Surface potential was calculated using the DELPHI web server (108) and visualized with CHIMERA (109). Crystallographic structures were deposited in the Protein Data Bank (PDB) under identification numbers 6T6E (HIV-1 Cter, NL4.3), 6T6I (HIV-1 Cter, subtype A2), and 6T6J (HIV-1 Cter, subtype A2, mutant NQKK).

**Analysis of the surface electrostatic potential of retroviral CTDs.** The structures and the sequences of the CTDs were extracted from HIV-1 A2 (PDB 6T6I) (present study), HIV-1 NL4.3 (PDB 6T6E) (present study), SIV (PDB 1C6V) (48), MVV (PDB 5LLJ) (49), RSV (PDB 1COM) (50), MMTV (PDB 5D7U) (51), MMLV (PDB 2M9U) (52), and PFV (PDB 4E7I) (53). The structure-based sequence alignment was performed using PROMALS3D (104). Structure superposition and root mean square deviation (RMSD) calculations were performed using secondary structure matching (SSM), superpose program (105) embedded in COOT (103). The sequence alignment representation was generated by ESPript (107). The surface



electrostatic potential was calculated using the DELPHI web server (108) and visualized with CHIMERA (109).

**Statistical tests.** All statistical analyses were performed on at least three independent experiments (transfection and transduction) using Prism 6 (GraphPad). For all functional tests, the values obtained for the chimeras were normalized using the values obtained for parental integrase A. Student's tests were used to evaluate whether the normalized mean values obtained with the chimeric and mutant integrases were significantly different from that obtained with the parental strain and/or between them.

**Data availability.** Crystallographic structures are available in PDB under identification numbers 6T6E, 6T6I, and 6T6J.

## ACKNOWLEDGMENTS

We are grateful to J. C. Plantier for providing HIV-1 strains of subtype B, CRF02, and group O, to C. Elefante for the construction of the p8.91-MB plasmid, to J. Batisse for providing control reagents, and to M. Lavigne, B. Maillot, and S. Marzi for helpful discussions. We thank R. Drillien (IGBMC) for suggestions about the manuscript. We thank V. Olieric and the staff of the Swiss Light Source synchrotron for help with data collection.

This work was supported by grants from Sidaction and the ANRS (grant ECTZ72120).

We also acknowledge the support by, and use of resources of, the French Infrastructure for Integrated Structural Biology (FRISBI ANR-10-INBS-05) and of Instruct-ERIC.

## REFERENCES

1. Pauza CD. 1990. Two bases are deleted from the termini of HIV-1 linear DNA during integrative recombination. *Virology* 179:886–889. [https://doi.org/10.1016/0042-6822\(90\)90161-j](https://doi.org/10.1016/0042-6822(90)90161-j).
2. Engelman A, Mizuuchi K, Craigie R. 1991. HIV-1 DNA integration—mechanism of viral-DNA cleavage and DNA strand transfer. *Cell* 67:1211–1221. [https://doi.org/10.1016/0092-8674\(91\)90297-c](https://doi.org/10.1016/0092-8674(91)90297-c).
3. Engelman A, Englund G, Orenstein JM, Martin MA, Craigie R. 1995. Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *J Virol* 69:2729–2736. <https://doi.org/10.1128/JVI.69.5.2729-2736.1995>.
4. Bukovsky A, Gottlinger H. 1996. Lack of integrase can markedly affect human immunodeficiency virus type 1 particle production in the presence of an active viral protease. *J Virol* 70:6820–6825. <https://doi.org/10.1128/JVI.70.10.6820-6825.1996>.
5. Hoyte AC, Jamin AV, Koneru PC, Kobe MJ, Larue RC, Fuchs JR, Engelman AN, Kvaratskhelia M. 2017. Resistance to pyridine-based inhibitor K116 reveals an unexpected role of integrase in HIV-1 Gag-Pol polyprotein proteolytic processing. *J Biol Chem* 292:19814–19825. <https://doi.org/10.1074/jbc.M117.816645>.
6. Kessl JJ, Kutluay SB, Townsend D, Rebensburg S, Slaughter A, Larue RC, Shkriabai N, Bakouche N, Fuchs JR, Bieniasz PD, Kvaratskhelia M. 2016. HIV-1 integrase binds the viral RNA genome and is essential during virion morphogenesis. *Cell* 166:1257–1268.e12. <https://doi.org/10.1016/j.cell.2016.07.044>.
7. Zhu K, Dobard C, Chow SA. 2004. Requirement for integrase during reverse transcription of human immunodeficiency virus type 1 and the effect of cysteine mutations of integrase on its interactions with reverse transcriptase. *J Virol* 78:5045–5055. <https://doi.org/10.1128/jvi.78.10.5045-5055.2004>.
8. Dobard CW, Briones MS, Chow SA. 2007. Molecular mechanisms by which human immunodeficiency virus type 1 integrase stimulates the early steps of reverse transcription. *J Virol* 81:10037–10046. <https://doi.org/10.1128/JVI.00519-07>.
9. Wilkinson TA, Januszyk K, Phillips ML, Tekeste SS, Zhang M, Miller JT, Le Grice SFJ, Clubb RT, Chow SA. 2009. Identifying and characterizing a functional HIV-1 reverse transcriptase-binding site on integrase. *J Biol Chem* 284:7931–7939. <https://doi.org/10.1074/jbc.M806241200>.
10. Emiliani S, Mousnier A, Busschots K, Maroun M, Van Maele B, Tempé D, Vandekerckhove L, Moisan F, Ben-Slama L, Witvrouw M, Christ F, Rain J-C, Dargemont C, Debyser Z, Benarous R. 2005. Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. *J Biol Chem* 280:25517–25523. <https://doi.org/10.1074/jbc.M501378200>.
11. Bukrinsky MI, Sharova N, Dempsey MP, Stanwick TL, Bukrinskaya AG, Haggerty S, Stevenson M. 1992. Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc Natl Acad Sci U S A* 89:6580–6584. <https://doi.org/10.1073/pnas.89.14.6580>.
12. Mattaj JW, Englmeier L. 1998. Nucleocytoplasmic transport: the soluble phase. *Annu Rev Biochem* 67:265–306. <https://doi.org/10.1146/annurev.biochem.67.1.265>.
13. Yamashita M, Emerman M. 2004. Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. *J Virol* 78:5670–5678. <https://doi.org/10.1128/JVI.78.11.5670-5678.2004>.
14. Yamashita M, Perez O, Hope TJ, Emerman M. 2007. Evidence for direct involvement of the capsid protein in HIV infection of nondividing cells. *PLoS Pathog* 3:e156. <https://doi.org/10.1371/journal.ppat.0030156>.
15. Di Nunzio F, Danckaert A, Fricke T, Perez P, Fernandez J, Perret E, Roux P, Shorte S, Charneau P, Diaz-Griffero F, Arhel NJ. 2012. Human nucleoporins promote HIV-1 docking at the nuclear pore, nuclear import and integration. *PLoS One* 7:e46037. <https://doi.org/10.1371/journal.pone.0046037>.
16. Di Nunzio F, Fricke T, Miccio A, Valle-Casuso JC, Perez P, Souque P, Rizzi E, Severgnini M, Mavilio F, Charneau P, Diaz-Griffero F. 2013. Nup153 and Nup98 bind the HIV-1 core and contribute to the early steps of HIV-1 replication. *Virology* 440:8–18. <https://doi.org/10.1016/j.virol.2013.02.008>.
17. Matreyek KA, Engelman A. 2011. The requirement for nucleoporin NUP153 during human immunodeficiency virus type 1 infection is determined by the viral capsid. *J Virol* 85:7818–7827. <https://doi.org/10.1128/JVI.00325-11>.
18. Krishnan L, Matreyek KA, Oztop I, Lee K, Tipper CH, Li X, Dar MJ, KewalRamani VN, Engelman A. 2010. The requirement for cellular transportin 3 (TNPO3 or TRN-SR2) during infection maps to human immunodeficiency virus type 1 capsid and not integrase. *J Virol* 84:397–406. <https://doi.org/10.1128/JVI.01899-09>.
19. Cribier A, Ségéral E, Delelis O, Parissi V, Simon A, Ruff M, Benarous R, Emiliani S. 2011. Mutations affecting interaction of integrase with TNPO3 do not prevent HIV-1 cDNA nuclear import. *Retrovirology* 8:104. <https://doi.org/10.1186/1742-4690-8-104>.
20. Gallay P, Hope T, Chin D, Trono D. 1997. HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. *Proc Natl Acad Sci U S A* 94:9825–9830. <https://doi.org/10.1073/pnas.94.18.9825>.
21. Bouyac-Bertoia M, Dvorin JD, Fouchier RAM, Jenkins Y, Meyer BE, Wu LI, Emerman M, Malim MH. 2001. HIV-1 infection requires a functional integrase NLS. *Mol Cell* 7:1025–1035. [https://doi.org/10.1016/S1097-2765\(01\)00240-4](https://doi.org/10.1016/S1097-2765(01)00240-4).
22. Hearsay AC, Jans DA. 2006. HIV-1 integrase is capable of targeting DNA to the nucleus via an Importin  $\alpha/\beta$ -dependent mechanism. *Biochem J* 398:475–484. <https://doi.org/10.1042/BJ20060466>.
23. Ao Z, Huang G, Yao H, Xu Z, Labine M, Cochrane AW, Yao X. 2007.



- Interaction of human immunodeficiency virus type 1 integrase with cellular nuclear import receptor importin 7 and its impact on viral replication. *J Biol Chem* 282:13456–13467. <https://doi.org/10.1074/jbc.M610546200>.
24. Ao Z, Jayappa KD, Wang B, Zheng Y, Kung S, Rassart E, Depping R, Kohler M, Cohen EA, Yao X. 2010. Importin  $\alpha$ 3 interacts with HIV-1 integrase and contributes to HIV-1 nuclear import and replication. *J Virol* 84:8650–8663. <https://doi.org/10.1128/JVI.00508-10>.
  25. Woodward CL, Prakobwanakit S, Mosessian S, Chow SA. 2009. Integrase interacts with nucleoporin NUP153 to mediate the nuclear import of human immunodeficiency virus type 1. *J Virol* 83:6522–6533. <https://doi.org/10.1128/JVI.02061-08>.
  26. Ao Z, Jayappa KD, Wang B, Zheng Y, Wang X, Peng J, Yao X. 2012. Contribution of host nucleoporin 62 in HIV-1 integrase chromatin association and viral DNA integration. *J Biol Chem* 287:10544–10555. <https://doi.org/10.1074/jbc.M111.317057>.
  27. Larue R, Gupta K, Wuensch C, Shkriabai N, Kessl JJ, Danhart E, Feng L, Taltyov O, Christ F, Van Duyn GD, Debyser Z, Foster MP, Kvaratskhelia M. 2012. Interaction of the HIV-1 intasome with transportin 3 protein (TNPO3 or TRN-SR2). *J Biol Chem* 287:34044–34058. <https://doi.org/10.1074/jbc.M112.384669>.
  28. De Houwer S, Demeulemeester J, Thys W, Rocha S, Dirix L, Gijbbers R, Christ F, Debyser Z. 2014. The HIV-1 integrase mutant R263A/K264A is 2-fold defective for TRN-SR2 binding and viral nuclear import. *J Biol Chem* 289:25351–25361. <https://doi.org/10.1074/jbc.M113.533281>.
  29. Christ F, Thys W, De Rijck J, Gijbbers R, Albanese A, Arosio D, Emiliani S, Rain J-C, Benarous R, Cereseto A, Debyser Z. 2008. Transportin-SR2 imports HIV into the nucleus. *Curr Biol* 18:1192–1202. <https://doi.org/10.1016/j.cub.2008.07.079>.
  30. Jayappa KD, Ao Z, Yang M, Wang J, Yao X. 2011. Identification of critical motifs within HIV-1 integrase required for importin  $\alpha$ 3 interaction and viral cDNA nuclear import. *J Mol Biol* 410:847–862. <https://doi.org/10.1016/j.jmb.2011.04.011>.
  31. Passos DO, Li M, Yang R, Rebensburg SV, Ghirlando R, Jeon Y, Shkriabai N, Kvaratskhelia M, Craigie R, Lyumkis D. 2017. Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science* 355:89–92. <https://doi.org/10.1126/science.aah5163>.
  32. Michel F, Crucifix C, Granger F, Eiler S, Mouscadet J-F, Korolev S, Agapkina J, Ziganshin R, Gottikh M, Nazabal A, Emiliani S, Benarous R, Moras D, Schultz P, Ruff M. 2009. Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/p75 cofactor. *EMBO J* 28:980–991. <https://doi.org/10.1038/emboj.2009.41>.
  33. Craigie R, Bushman FD. 2012. HIV DNA integration. *Cold Spring Harb Perspect Med* 2:a006890. <https://doi.org/10.1101/cshperspect.a006890>.
  34. Delelis O, Carayon K, Saïb A, Deprez E, Mouscadet J-F. 2008. Integrase and integration: biochemical activities of HIV-1 integrase. *Retrovirology* 5:114. <https://doi.org/10.1186/1742-4690-5-114>.
  35. Zheng R, Jenkins TM, Craigie R. 1996. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc Natl Acad Sci U S A* 93:13659–13664. <https://doi.org/10.1073/pnas.93.24.13659>.
  36. Eijkelenboom AP, van den Ent FM, Vos A, Doreleijers JF, Hård K, Tullius TD, Plasterk RH, Kaptein R, Boelens R. 1997. The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc. *Curr Biol* 7:739–746. [https://doi.org/10.1016/S0960-9822\(06\)00332-0](https://doi.org/10.1016/S0960-9822(06)00332-0).
  37. Busschots K, Vercammen J, Emiliani S, Benarous R, Engelborghs Y, Christ F, Debyser Z. 2005. The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J Biol Chem* 280:17841–17847. <https://doi.org/10.1074/jbc.M411681200>.
  38. Heuer TS, Brown PO. 1997. Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. *Biochemistry* 36:10655–10665. <https://doi.org/10.1021/bi970782h>.
  39. Esposito D, Craigie R. 1998. Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J* 17:5832–5843. <https://doi.org/10.1093/emboj/17.19.5832>.
  40. Chen A, Weber IT, Harrison RW, Leis J. 2006. Identification of amino acids in HIV-1 and avian sarcoma virus integrase subsites required for specific recognition of the long terminal repeat ends. *J Biol Chem* 281:4173–4182. <https://doi.org/10.1074/jbc.M510628200>.
  41. Engelman A, Hickman AB, Craigie R. 1994. The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *J Virol* 68:5911–5917. <https://doi.org/10.1128/JVI.68.9.5911-5917.1994>.
  42. Lutzke RA, Vink C, Plasterk RH. 1994. Characterization of the minimal DNA-binding domain of the HIV integrase protein. *Nucleic Acids Res* 22:4125–4131. <https://doi.org/10.1093/nar/22.20.4125>.
  43. Cannon PM, Byles ED, Kingsman SM, Kingsman AJ. 1996. Conserved sequences in the carboxyl terminus of integrase that are essential for human immunodeficiency virus type 1 replication. *J Virol* 70:651–657. <https://doi.org/10.1128/JVI.70.1.651-657.1996>.
  44. Larder BA, Purifoy DJ, Powell KL, Darby G. 1987. Site-specific mutagenesis of AIDS virus reverse transcriptase. *Nature* 327:716–717. <https://doi.org/10.1038/327716a0>.
  45. Mandal D, Feng Z, Stoltzfus CM. 2008. Gag-processing defect of human immunodeficiency virus type 1 integrase E246 and G247 mutants is caused by activation of an overlapping 5' splice site. *J Virol* 82:1600–1604. <https://doi.org/10.1128/JVI.02295-07>.
  46. Hemelaar J, Gouws E, Ghys PD, Osmanov S, WHO-UNAIDS Network for HIV Isolation and Characterisation. 2011. Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS* 25:679–689. <https://doi.org/10.1097/QAD.0b013e328342ff93>.
  47. Sloan RD, Wainberg MA. 2011. The role of unintegrated DNA in HIV infection. *Retrovirology* 8:52. <https://doi.org/10.1186/1742-4690-8-52>.
  48. Chen Z, Yan Y, Munshi S, Li Y, Zugay-Murphy J, Xu B, Witmer M, Felock P, Wolfe A, Sardana V, Emini EA, Hazuda D, Luo LC. 2000. X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50–293): an initial glance of the viral DNA binding platform. *J Mol Biol* 296:521–533. <https://doi.org/10.1006/jmbi.1999.3451>.
  49. Ballandras-Colas A, Maskell DP, Serrao E, Locke J, Swuec P, Jónsson SR, Kotecha A, Cook NJ, Pye VE, Taylor IA, Andrésdóttir V, Engelman AN, Costa A, Cherepanov P. 2017. A supramolecular assembly mediates lentiviral DNA integration. *Science* 355:93–95. <https://doi.org/10.1126/science.aah7002>.
  50. Yang ZN, Mueser TC, Bushman FD, Hyde CC. 2000. Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J Mol Biol* 296:535–548. <https://doi.org/10.1006/jmbi.1999.3463>.
  51. Ballandras-Colas A, Brown M, Cook NJ, Dewdney TG, Demeler B, Cherepanov P, Lyumkis D, Engelman AN. 2016. Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* 530:358–361. <https://doi.org/10.1038/nature16955>.
  52. Aiyer S, Swapna GVT, Malani N, Aramini JM, Schneider WM, Plumb MR, Ghanem M, Larue RC, Sharma A, Studamire B, Kvaratskhelia M, Bushman FD, Montelione GT, Roth MJ. 2014. Altering murine leukemia virus integration through disruption of the integrase and BET protein family interaction. *Nucleic Acids Res* 42:5917–5928. <https://doi.org/10.1093/nar/gku175>.
  53. Hare S, Maertens GN, Cherepanov P. 2012. 3'-Processing and strand transfer catalysed by retroviral integrase in crystallo. *EMBO J* 31:3020–3028. <https://doi.org/10.1038/emboj.2012.118>.
  54. Luscombe NM, Laskowski RA, Thornton JM. 2001. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res* 29:2860–2874. <https://doi.org/10.1093/nar/29.13.2860>.
  55. Zhou H-X, Pang X. 2018. Electrostatic interactions in protein structure, folding, binding, and condensation. *Chem Rev* 118:1691–1741. <https://doi.org/10.1021/acs.chemrev.7b00305>.
  56. Karlin S, Zhu ZY. 1996. Characterizations of diverse residue clusters in protein three-dimensional structures. *Proc Natl Acad Sci U S A* 93:8344–8349. <https://doi.org/10.1073/pnas.93.16.8344>.
  57. Karlin S, Brendel V. 1988. Charge configurations in viral proteins. *Proc Natl Acad Sci U S A* 85:9396–9400. <https://doi.org/10.1073/pnas.85.24.9396>.
  58. Parker MS, Balasubramaniam A, Parker SL. 2012. On the segregation of protein ionic residues by charge type. *Amino Acids* 43:2231–2247. <https://doi.org/10.1007/s00726-012-1418-4>.
  59. Kharrat N, Belmabrouk S, Abdelhedi R, Benmarzoug R, Assidi M, Qahtani AI, Rebai A. 2016. Screening for clusters of charge in human virus proteomes. *BMC Genomics* 17:758. <https://doi.org/10.1186/s12864-016-3086-3>.
  60. Dostálková A, Kaufman F, Křížová I, Kultová A, Strohalmová K, Hadrová R, Ruml T, Rumlová M. 2018. Mutations in the basic region of the Mason-Pfizer monkey virus nucleocapsid protein affect reverse transcription, genomic RNA packaging, and the virus assembly site. *J Virol* 92:5439. <https://doi.org/10.1128/JVI.00106-18>.

61. Hamann MV, Müllers E, Reh J, Stanke N, Effantin G, Weissenhorn W, Lindemann D. 2014. The cooperative function of arginine residues in the prototype foamy virus Gag C-terminus mediates viral and cellular RNA encapsidation. *Retrovirology* 11:87–17. 6 ed. <https://doi.org/10.1186/s12977-014-0087-7>.
62. Heyrana KJ, Goh BC, Perilla JR, Nguyen T-LN, England MR, Bewley MC, Schulten K, Craven RC. 2016. Contributions of charged residues in structurally dynamic capsid surface loops to Rous sarcoma virus assembly. *J Virol* 90:5700–5714. <https://doi.org/10.1128/JVI.00378-16>.
63. Yenamandra SP, Sompallae R, Klein G, Kashuba E. 2009. Comparative analysis of the Epstein-Barr virus encoded nuclear proteins of EBNA-3 family. *Comput Biol Med* 39:1036–1042. <https://doi.org/10.1016/j.combiomed.2009.08.006>.
64. Roller RJ, Bjerke SL, Haugo AC, Hanson S. 2010. Analysis of a charge cluster mutation of herpes simplex virus type 1 UL34 and its extragenic suppressor suggests a novel interaction between pUL34 and pUL31 that is necessary for membrane curvature around capsids. *J Virol* 84:3921–3934. <https://doi.org/10.1128/JVI.01638-09>.
65. Zhang J, Webb DM. 2004. Rapid evolution of primate antiviral enzyme APOBEC3G. *Hum Mol Genet* 13:1785–1791. <https://doi.org/10.1093/hmg/ddh183>.
66. Vasudev PG, Banerjee M, Ramakrishnan C, Balam P. 2012. Asparagine and glutamine differ in their propensities to form specific side chain-backbone hydrogen bonded motifs in proteins. *Proteins* 80:991–1002. <https://doi.org/10.1002/prot.24001>.
67. Weichenberger CX, Sippl MJ. 2006. Self-consistent assignment of asparagine and glutamine amide rotamers in protein crystal structures. *Structure* 14:967–972. <https://doi.org/10.1016/j.str.2006.04.002>.
68. Métifiot M, Johnson BC, Kiselev E, Marler L, Zhao XZ, Burke TR, Marchand C, Hughes SH, Pommier Y. 2016. Selectivity for strand-transfer over 3'-processing and susceptibility to clinical resistance of HIV-1 integrase inhibitors are driven by key enzyme-DNA interactions in the active site. *Nucleic Acids Res* 44:6896–6906. <https://doi.org/10.1093/nar/gkw592>.
69. Johnson AA, Santos W, Pais GCG, Marchand C, Amin R, Burke TR, Verdine G, Pommier Y. 2006. Integration requires a specific interaction of the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop. *J Biol Chem* 281:461–467. <https://doi.org/10.1074/jbc.M511348200>.
70. Guiot E, Carayon K, Delelis O, Simon F, Tauc P, Zubin E, Gottikh M, Mouscadet J-F, Brochon J-C, Deprez E. 2006. Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity. *J Biol Chem* 281:22707–22719. <https://doi.org/10.1074/jbc.M602198200>.
71. Hehl EA, Joshi P, Kalpana GV, Prasad VR. 2004. Interaction between human immunodeficiency virus type 1 reverse transcriptase and integrase proteins. *J Virol* 78:5056–5067. <https://doi.org/10.1128/jvi.78.10.5056-5067.2004>.
72. Ao Z, Fowke KR, Cohen EA, Yao X. 2005. Contribution of the C-terminal tri-lysine regions of human immunodeficiency virus type 1 integrase for efficient reverse transcription and viral DNA nuclear import. *Retrovirology* 2:62. <https://doi.org/10.1186/1742-4690-2-62>.
73. Williams KL, Zhang Y, Shkriabai N, Karki RG, Nicklaus MC, Kotrikadze N, Hess S, Le Grice SFJ, Craigie R, Pathak VK, Kvaratskhelia M. 2005. Mass spectrometric analysis of the HIV-1 integrase-pyridoxal 5'-phosphate complex reveals a new binding site for a nucleotide inhibitor. *J Biol Chem* 280:7949–7955. <https://doi.org/10.1074/jbc.M413579200>.
74. Tekeste SS, Wilkinson TA, Weiner EM, Xu X, Miller JT, Le Grice SFJ, Clubb RT, Chow SA. 2015. Interaction between reverse transcriptase and integrase is required for reverse transcription during HIV-1 replication. *J Virol* 89:12058–12069. <https://doi.org/10.1128/JVI.01471-15>.
75. Dar MJ, Monel B, Krishnan L, Shun M-C, Di Nunzio F, Helland DE, Engelman A. 2009. Biochemical and virological analysis of the 18-residue C-terminal tail of HIV-1 integrase. *Retrovirology* 6:94. <https://doi.org/10.1186/1742-4690-6-94>.
76. Mohammed KD, Topper MB, Muesing MA. 2011. Sequential deletion of the integrase (Gag-Pol) carboxyl terminus reveals distinct phenotypic classes of defective HIV-1. *J Virol* 85:4654–4666. <https://doi.org/10.1128/JVI.02374-10>.
77. Lu R, Ghory HZ, Engelman A. 2005. Genetic analyses of conserved residues in the carboxyl-terminal domain of human immunodeficiency virus type 1 integrase. *J Virol* 79:10356–10368. <https://doi.org/10.1128/JVI.79.16.10356-10368.2005>.
78. Cereseto A, Manganaro L, Gutierrez MI, Terreni M, Fittipaldi A, Lusic M, Marcello A, Giacca M. 2005. Acetylation of HIV-1 integrase by p300 regulates viral integration. *EMBO J* 24:3070–3081. <https://doi.org/10.1038/sj.emboj.7600770>.
79. Terreni M, Valentini P, Liverani V, Gutierrez MI, Di Primio C, Di Fenza A, Tozzini V, Allouch A, Albanese A, Giacca M, Cereseto A. 2010. GCN5-dependent acetylation of HIV-1 integrase enhances viral integration. *Retrovirology* 7:18. <https://doi.org/10.1186/1742-4690-7-18>.
80. Topper M, Luo Y, Zhadina M, Mohammed K, Smith L, Muesing MA. 2007. Posttranslational acetylation of the human immunodeficiency virus type 1 integrase carboxyl-terminal domain is dispensable for viral replication. *J Virol* 81:3012–3017. <https://doi.org/10.1128/JVI.02257-06>.
81. Zufferey R, Nagy D, Mandel RJ, Naldini L, Trono D. 1997. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat Biotechnol* 15:871–875. <https://doi.org/10.1038/nbt0997-871>.
82. Rossolillo P, Winter F, Simon-Loriere E, Gallois-Montbrun S, Negroni M. 2012. Retrovirology: HIV-driven evolution of cellular genes and improvement of anticancer drug activation. *PLoS Genet* 8:e1002904. <https://doi.org/10.1371/journal.pgen.1002904>.
83. Adachi A, Gendelman HE, Koenig S, Folks T, Willey R, Rabson A, Martin MA. 1986. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J Virol* 59:284–291. <https://doi.org/10.1128/JVI.59.2.284-291.1986>.
84. Gasser R, Hamoudi M, Pellicciotta M, Zhou Z, Visdeloup C, Colin P, Braibant M, Lagane B, Negroni M. 2016. Buffering deleterious polymorphisms in highly constrained parts of HIV-1 envelope by flexible regions. *Retrovirology* 13:50. <https://doi.org/10.1186/s12977-016-0285-6>.
85. Charneau P, Mirambeau G, Roux P, Paulous S, Buc H, Clavel F. 1994. HIV-1 reverse transcription. A termination step at the center of the genome. *J Mol Biol* 241:651–662. <https://doi.org/10.1006/jmbi.1994.1542>.
86. Rosen CA, Sodroski JG, Campbell K, Haseltine WA. 1986. Construction of recombinant murine retroviruses that express the human T-cell leukemia virus type II and human T-cell lymphotropic virus type III trans activator genes. *J Virol* 57:379–384. <https://doi.org/10.1128/JVI.57.1.379-384.1986>.
87. Foley G, Lazarus H, Farber S, Uzman B, Boone B, McCarthy R. 1965. Continuous culture of human lymphoblasts from peripheral blood of a child with acute leukemia. *Cancer* 18:522–529. [https://doi.org/10.1002/1097-0142\(196504\)18:4<522::AID-CNCR2820180418>3.0.CO;2-J](https://doi.org/10.1002/1097-0142(196504)18:4<522::AID-CNCR2820180418>3.0.CO;2-J).
88. Nara PL, Fischinger PJ. 1988. Quantitative infectivity assay for HIV-1 and-2. *Nature* 332:469–470. <https://doi.org/10.1038/332469a0>.
89. Nara PL, Hatch WC, Dunlop NM, Robey WG, Arthur LO, Gonda MA, Fischinger PJ. 1987. Simple, rapid, quantitative, syncytium-forming microassay for the detection of human immunodeficiency virus neutralizing antibody. *AIDS Res Hum Retroviruses* 3:283–302. <https://doi.org/10.1089/aid.1987.3.283>.
90. Naldini L, Blömer U, Gallay P, Ory D, Mulligan R, Gage FH, Verma IM, Trono D. 1996. In vivo gene delivery and stable transduction of non-dividing cells by a lentiviral vector. *Science* 272:263–267. <https://doi.org/10.1126/science.272.5259.263>.
91. Vozzolo L, Loh B, Gane PJ, Tribak M, Zhou L, Anderson I, Nyakatura E, Jenner RG, Selwood D, Fassati A. 2010. Gyrase B inhibitor impairs HIV-1 replication by targeting Hsp90 and the capsid protein. *J Biol Chem* 285:39314–39328. <https://doi.org/10.1074/jbc.M110.155275>.
92. De Iaco A, Santoni F, Vannier A, Guipponi M, Antonarakis S, Luban J. 2013. TNPO3 protects HIV-1 replication from CPSF6-mediated capsid stabilization in the host cell cytoplasm. *Retrovirology* 10:20. <https://doi.org/10.1186/1742-4690-10-20>.
93. Sarzotti-Kelsoe M, Bailer RT, Turk E, Lin C-L, Bilska M, Greene KM, Gao H, Todd CA, Ozaki DA, Seaman MS, Mascola JR, Montefiori DC. 2014. Optimization and validation of the TZM-bl assay for standardized assessments of neutralizing antibodies against HIV-1. *J Immunol Methods* 409:131–146. <https://doi.org/10.1016/j.jim.2013.11.022>.
94. Lederle A, Su B, Holl V, Penichon J, Schmidt S, Decoville T, Laumond G, Moog C. 2014. Neutralizing antibodies inhibit HIV-1 infection of plasmacytoid dendritic cells by an FcγRIIa independent mechanism and do not diminish cytokines production. *Sci Rep* 4:5845. <https://doi.org/10.1038/srep05845>.
95. Kabsch W. 2010. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* 66:133–144. <https://doi.org/10.1107/S0907444909047374>.
96. Kabsch W. 2010. XDS. *Acta Crystallogr D Biol Crystallogr* 66:125–132. <https://doi.org/10.1107/S0907444909047337>.
97. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read

- RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr* 40: 658–674. <https://doi.org/10.1107/S0021889807021206>.
98. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66: 213–221. <https://doi.org/10.1107/S0907444909052925>.
  99. Eijkelenboom APAM, Sprangers R, Hård K, Puras Lutzke RA, Plasterk RHA, Boelens R, Kaptein R. 1999. Refined solution structure of the C-terminal DNA-binding domain of human immunovirus-1 integrase. *Proteins* 36: 556–564. [https://doi.org/10.1002/\(SICI\)1097-0134\(19990901\)36:4<556::AID-PROT18>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1097-0134(19990901)36:4<556::AID-PROT18>3.0.CO;2-6).
  100. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Adams PD, Read RJ, Zwart PH, Hung L-W. 2008. Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallogr D Biol Crystallogr* 64:515–524. <https://doi.org/10.1107/S0907444908004319>.
  101. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung L-W, Read RJ, Adams PD. 2008. Iterative model building, structure refinement and density modification with the PHENIX Auto-Build wizard. *Acta Crystallogr D Biol Crystallogr* 64:61–69. <https://doi.org/10.1107/S090744490705024X>.
  102. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD. 2012. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68:352–367. <https://doi.org/10.1107/S0907444912001308>.
  103. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501. <https://doi.org/10.1107/S0907444910007493>.
  104. Pei J, Kim B-H, Grishin NV. 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* 36: 2295–2300. <https://doi.org/10.1093/nar/gkn072>.
  105. Krissinel E, Henrick K. 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60:2256–2268. <https://doi.org/10.1107/S0907444904026460>.
  106. Hough MA, Wilson KS. 2018. From crystal to structure with CCP4. *Acta Crystallogr D Struct Biol* 74:67–67. <https://doi.org/10.1107/S2059798317017557>.
  107. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42:W320–4. <https://doi.org/10.1093/nar/gku316>.
  108. Sarkar S, Witham S, Zhang J, Zhenirovskyy M, Rocchia W, Alexov E. 2013. DelPhi Web Server: a comprehensive online suite for electrostatic calculations of biological macromolecules and their complexes. *Commun Comput Phys* 13:269–284. <https://doi.org/10.4208/cicp.300611.201011s>.
  109. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera: a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. <https://doi.org/10.1002/jcc.20084>.